LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH Ollscoil Átha Cliath

TRINITY COLLEGE LIBRARY DUBLIN The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Trust as a Form of Defeasible Reasoning

Computing the Defeasible Expertise of Trust using Defeasible Reasoning and Generic Patterns

Pierpaolo Dondio

A thesis submitted to the University of Dublin, Trinity College in partial fulfillment of the requirements for the degree of Doctor of Philosophy

October 2008

Declaration

I, the undersigned, declare that this work has not previously been submitted to this or any other University, and that unless otherwise stated, it is entirely my own work.

Pierpaolo Dondio

Dated: 29th October 2008

TRINITY COLLEGE

2 7 JUL 2011

LIBRARY DUBLIN

7 H68/8

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Pierpaolo Dondio

Dated: 29th October 2008

Per illy

Acknowledgements

First and foremost, I would like to thank my supervisor Stephen Barrett for his guidance, support, diligence and mentoring on the long journey. He has always been patient and prodding when I needed it. Moreover, I would like to thank Stephen for his constant optimism and enthusiasm, saving this PhD at its early stages. You always gave me confidence and encouragement, thanks again. Moreover, thanks for becoming a nice friend of mine.

I would like to thank Dr. Jean-Marc Seigneur, for having introduced me to Computational Trust and academia. He is surely one of the most trustworthy people I have ever met. A special thank to Prof. E. Mullis, for the review of the statistical section of this thesis. The other main acknowledgements must go to the present and past people at DSG.

Finally, thanks to my wife Ilaria. This thesis is dedicated to you for all your support and unconditional trust you had in myself, and for being my special mate during this long journey.

Pierpaolo Dondio University of Dublin, Trinity College Dublin, 29th October 2008

Summary

Computational models of Trust have recently emerged as a way to exploit the human notion of trust in open and collaborative environments.

This thesis provides a novel computational model of trust based on defeasible reasoning. Our starting point is an analysis of the nature of trust, that appears as a form of reasoning where arguments used are defeasible rather than deductive, subject to be defeated or strengthened in light of new evidence or conflicting arguments. The research issue investigated is the following; if trust is a form of defeasible reasoning, the application of defeasible reasoning computational techniques should improve the quality of trust computation.

Defeasible reasoning reaches conclusions by considering the logical consistency of arguments rather than simply aggregating and merging them. During the defeasible trust-based reasoning, weak arguments are defeated by other arguments, reducing the number of false predictions, while strong and plausible arguments are strengthened, increasing the number of correct predictions.

In order to build a trust model based on defeasible reasoning, our theoretical contribution, the following tasks have to be accomplished: (i) definition of those recurrent arguments that compose a trust decision, (ii) the study of their defeasibility and (iii) the definition of a semantic for combining and aggregating different pieces of evidence.

In order to accomplish task 1, we investigated the current landscape of computational models of trust in order to identify the recurrent patterns –if any- underlying a trust-based decisions. Moreover, we investigated the state-of-the-art in social science and cognitive models of trust in order to identify new mechanisms accepted by researchers but not yet computationally investigated. We produced a new version of them as *defeasible* arguments by investigating assumptions underlying each mechanism, identifying possible supporters or defeaters arguments and a way to assess their plausibility. Finally, we identified a set of mutual relationships among arguments and we defined a semantic to combine contrasting and supporting arguments into a set of justifiable conclusions.

In order to evaluate our model, we defined an evaluation strategy based on the

comparative analysis of our computation with a reliable and independent trust metric, selecting only applications where the hypothesis of *common understanding* can be satisfied.

We evaluate the model in the context of a large online web community and the Wikipedia project, using large sets of data describing users' activity in these applications. Our computation shows how the introduction of defeasibility increases the quality of our computation by 10% to 50%. On an absolute scale, results obtained show a very high degree of precision of 80% to 90% rate of good predictions. Our evaluation shows also the validity of novel trust mechanisms introduced, like the study of persistency, activity and stability, that allow computing trust in an alternative way other than the classical reputation and indirect experience mechanisms. The defeasible nature of trust mechanisms, including those usually regarded as *objective*, is shown. The defeasible computation performed produces interesting features: it could be application-contained; it extends the set of evidence used in trust. We showed how these features are made possible by the introduction of defeasibility.

The main contributions of the thesis are therefore the following: the definition of a trust model based on defeasible reasoning – our theoretical contribution -, encompassing the definition of trust arguments, their defeasibility study and a computable semantic for combining them; and a new and alternative trust computation showing a high degree of accuracy.

Relevant Author's publications

[Don07e] P. Dondio, S. Barrett, Presumptive Selections of Trust Evidences. AAMAS 2007, 6th joint international conference on Multi-Agents and Autonomous Systems, Hawaii, 2007, USA

[Don07c] P. Dondio, S. Barrett, Computational Trust in Web content quality *Informatica Journal*, N. 31, pages. 151-160, June 2007

[Sei07] JM. Seigneur, P. Dondio. Trust in self-organizing systems, chapter of the book *Self-organizing systems* edited by Springer. To be published December 2009

[Lon07] L. Longo, P. Dondio, S. Barrett, Temporal Factors to evaluate trustworthiness of virtual identities, proceedings of *IEEE SECOVAL 2007, Third International Workshop on the Value of Security through Collaboration, SECURECOM 2007*, Nice, France, September 2007

[Don07a] P. Dondio, E. Manzo, S. Barrett, Applied Computational Trust in Utilities Management a Case Study on The Town Council of Cava dei Tirreni, in *Trust Management, proceedings of IFIPTM, the first joint iTrust and PST Conferences on Privacy, Trust Management and Security*, Springer, July 2007.

[Don07b] P. Dondio, L. Longo. A translation Mechanism for Recommendations. *Proceedings of the 2nd IFIP joint conference on Trust Management*, Trondheim, Norway, June, 2008

[Don07d] P. Dondio, S. Barrett. Application-Contained Trust Calculation: a non-Invasive Approach Based on Presumptive Reasoning and Intuitive Trust. *UbiSafe*, *IEEE Symposium on Ubisafe Computing*, Niagara Falls, 2007, Canada

[Don06a] P. Dondio et al. Extracting Trust from Domain Analysis: a Study Case on the Wikipedia Project, *IEEE Automatic Trusted Computing Conference*, LNCS, 2006, Wuang, China

[Don06b] P. Dondio, S. Barrett, S. Weber: Calculating the Trustworthiness of a Wikipedia Article Using DANTE Methodology, *Proceedings of IADIS eSociety conference* 2006, Dublin

Content

| Cha | apter 1 | Introduction ust, from Social to Computer Science | | .18 |
|----------------|----------------|---|----|-----|
| 1.2 | | r Starting Point: trust as a defeasible phenomenon | | |
| 1.3 | | search Issue | | |
| 1.4 | | del Quick Overview | | |
| 1.5 | | aluation Strategy: comparative evaluation of Trust Metrics | | |
| 1.6 | | esis Contributions | | |
| 1.7 | | esis Structure | | |
| | | | | |
| Cha 2.1 | apter 2 Tru | St in the social sciences | | 40 |
| 2 | .1.1 | Romano's ten defining characteristics of trust. | 40 | |
| | Subje | ectivity vs. objectivity of Trust | 44 | |
| 2 | .1.2 | Definition of Trust | 45 | |
| 2.2 | Tru | st in computational models: definitions and properties | | .46 |
| 2 | .2.2 Fc | ormal Model of Trust: trust function, values and properties | 47 | |
| | 2.2.2 | .1 The Trust Function | 47 | |
| | 2.2.2 | 2 Trust Representation | 50 | |
| | 2.2.2 | .3 Properties of the Trust function | 54 | |
| 2.3 | Co | mputational Trust Solution: components and methodology | | 58 |
| 2.4 | Co | mputational models: how to compute trust | | 60 |
| 2 | .4.1 | Past Outcome using Direct Experience | 61 | |
| 2 | .4.2 | The Theory of Probability as a Computational Tool in Trust | 62 | |
| 2 | .4.3 | Indirect Experience | 64 | |
| 2 | .4.4 | Collaborative filtering, Similarity and Categorization | 66 | |
| 2 | .4.5 | Game Theory: the utility game | 67 | |
| 2 | .4.6 | Risk | 67 | |
| 2 | .4.7 | Cognitive models and Trust Ingredients | 68 | |
| T | he ing | redients of Trust | 68 | |
| 2 | .4.8 | Computing trust over domain elements: Monitoring and Heuristics | 69 | |
| 2.5 | Th | e problem of evidence selection | | 71 |
| 2 | .5.1 | Type and location of evidence | 71 | |
| | 2.5.1 | 1 Ad-hoc evidence in a dedicated trust infrastructure | 71 | |

| 2.5.1.2 Using feedback as evidence: humans' ratings, outcomes, trust values, | 72 |
|---|--------------------------|
| 2.5.1.3 Application elements as trust evidence: with or without trust | 73 |
| Conclusions | |
| | |
| Chapter 3 Presumptive and Defeasible Reasoning 3.1 Argumentation Theory and Non-monotonic Reasoning | |
| 3.1.1 Non-monotonic reasoning | |
| 3.1.2 Explanatory Reasoning and Abduction | 81 |
| 3.1.3 Default Logic and Closed World Assumption | 83 |
| 3.1.4 Closed World Assumption | |
| 3.2 Presumptive Reasoning | |
| 3.2.1 What is presumptive reasoning? | 85 |
| 3.2.2 The dynamic of Presumptive Reasoning and the Burden of Proof | 87 |
| 3.2.3 The presumptive Argumentation Scheme and the critical Questions | |
| 3.4 Defeasible Reasoning and the problem of combining arguments | |
| 3.3.1 Pollock's Defeasible Reasoning | 94 |
| 3.3.1.1 Various Degree of Justifications | 99 |
| 3.4 Trust and Defeasible Reasoning | |
| | |
| Conclusions | 1 |
| | 1 |
| 2.2.2.2 Trust Representation | |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning | 1 |
| 2.2.2.2 Trust Representation | 108 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning | 108 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning. Trust as reasoning. Trust as a presumption. Trust as a Presumptive and Defeasible Reasoning. | 108 109 110 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning | 108109110 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning 4.2 Trust as defeasible reasoning: the inference graph | 108109110 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning 4.2 Trust as defeasible reasoning: the inference graph 4.3 The (informal) notion of Trust Scheme | 108109110 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning. Trust as reasoning. Trust as a presumption. Trust as a Presumptive and Defeasible Reasoning. 4.2 Trust as defeasible reasoning: the inference graph. 4.2.1 Notation. 4.3 The (informal) notion of Trust Scheme. | 108109110114116 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning 4.2 Trust as defeasible reasoning: the inference graph 4.3.1 Notation 4.3.1 Trust Scheme and critical questions | 108109110114116118 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning | 108109110114116118120 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning 4.2 Trust as defeasible reasoning: the inference graph 4.2.1 Notation 4.3 The (informal) notion of Trust Scheme 4.3.1 Trust Scheme and critical questions 4.3.2 Validity of a Trust scheme 4.3.3 Source of Trust Scheme 4.3.4 Example of Trust Scheme | |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning | 108109110114116118120120 |
| Chapter 4 A model of Trust as Defeasible Reasoning 4.1 Trust is a form of Defeasible Reasoning Trust as reasoning Trust as a presumption Trust as a Presumptive and Defeasible Reasoning 4.2 Trust as defeasible reasoning: the inference graph 4.2.1 Notation 4.3 The (informal) notion of Trust Scheme 4.3.1 Trust Scheme and critical questions 4.3.2 Validity of a Trust scheme 4.3.3 Source of Trust Scheme 4.3.4 Example of Trust Scheme 4.4 Overall design of the method, operational description | 108109110114116118120120 |

| 4.4.4 | Stage 4: Exogenous factors and final decision | 131 | |
|-----------|--|-----|---------|
| 4.5 For | rmal Definitions of Trust Scheme | | .133 |
| 4.5.1 | Trust-based reasoning Graph | 133 | |
| 4.5.2 | Trust scheme definition | 134 | |
| 4.6 Ser | mantic | | .136 |
| Rule 1: | Strength of a conclusion – function F _c | 136 | |
| Rule 2: | Activation of links. | 138 | |
| Rule 3: | Accrual of reasons | 139 | |
| Rule 4: | Role of defeaters | 140 | |
| 4.6.1 | Note on the computation | 141 | |
| 4.6.1 | .1 Argumentation and Computation | 142 | |
| 4.7 Mo | odel's Contribution | | .144 |
| Conclusio | ns | | .145 |
| | | | |
| Chapter 5 | 5 Implementing the Model nking-based Computation | | 147 |
| 5.1.2 | Positive and Negative evidence | | . 1 1 / |
| | feasible Trust Scheme | | 150 |
| 5.2.1 | Time-based Trust Schemes | | |
| 5.2.2 | Trust Scheme based on Information Sharing and Social Role | | |
| | ation Sharing | | |
| | Schemes linked to Social Role | | |
| 5.2.3 | Trust Schemes linked to Activity | | |
| 5.2.4 | Trust Scheme linked to outcomes | | |
| 5.2.5 | Trust Scheme based on prejudices and grouping | | |
| 5.2.6 | Game theoretical/Cognitive Trust Scheme and Risk | | |
| 5.2.7 | Trust scheme summary | | |
| 5.3 Mutua | l relationships among Trust Schemes | | .161 |
| | ombining trust arguments: mutual relationships and argumentation | | |
| 5.3.2 | Mutual relationships among trust schemes | | |
| 5.4 Seman | ntic | | .166 |
| 5.4.1 | Strength of the conclusion | | |
| 5.4.2 | Function of Support | | |
| 5.4.3 | Function of attack: defeating arguments | | |
| 5.4.4 | The accrual of reasons: the Aggregation of Different Rankings | | |

| 5.4.5 Final Aggregation | 174 |
|--|-----|
| Conclusions | 175 |
| | |
| Chapter 6 Evaluation 6.1 Aspects under evaluation | |
| 6.1.1 List of Experiments | 179 |
| 6.2 Analysis of the Experiments | 180 |
| 6.2.1 FinanzaOnline.it, trust-based mining of a large on-line Community | 180 |
| Investigation of an existing trust-ranking | 180 |
| Application model | 182 |
| 6.2.2 Experiment I – Full application of the Method | 184 |
| Evidence Selection phase | 184 |
| Data and Indicators Selected for Trust Scheme | 186 |
| Time-based Trust Schemes | |
| Longevity | |
| Persistency | |
| Trust Scheme Stability | |
| Activity-based Trust Scheme | |
| Competence/Pertinence | 192 |
| Social Role Trust Scheme | |
| (Connectivity and) Authority | 194 |
| Info provisioning. | 197 |
| Visibility | 197 |
| Past-Outcomes | 198 |
| Trust Scheme based on statistic and grouping | 198 |
| Recommendation System | 199 |
| Defeasible Argumentation Phase | 199 |
| 6.2.2.1 Results evaluation | 199 |
| 6.2.3 Exp. II – Full application of the method using a limited set of evidence | 203 |
| Argumentation with the basic formula | 205 |
| Final comment. | 206 |
| 6.2.4 Exp. III – The Defeasibility of Past-Outcome Trust Scheme | |
| Analysis of results | 209 |
| 6.2.5 Exp. IV – Info Provisioning Trust Scheme Analysis | |
| 6.3 The case of Wikipedia: a trust reasoning for wiki-based applications | 21 |

| | Dataset used and plausibility of the comparative evaluation | 213 | |
|--|---|-----|-------|
| 6.3.1 Exp. V – Trustworthiness of Wikipedia Articles | | | |
| Wikipedia Application Model | | | |
| Evidence Selection, trust scheme, critical question | | | |
| | Activity | 215 | |
| | Stability | 217 | |
| | Pluralism | 218 | |
| | Longevity | 219 | |
| | Similarity/Categorization | 219 | |
| | Standard Compliance | 220 | |
| | Authorship | 221 | |
| 6.3 | 3.1.1 Results | 221 | |
| 6.3 | 3.2 Exp. VI – Trustworthiness of Wikipedia Authors | 224 | |
| | Analysis of the Schemes used | 225 | |
| | Longevity | 225 | |
| | Persistency and Regularity | 225 | |
| | Activity | 226 | |
| | Info provisioning | 226 | |
| Accessibility | | 226 | |
| | Past-Outcomes | 227 | |
| | Results | 227 | |
| 6.3 | 3.3 Exp. VII – Evaluation of a PageRank-based heuristic for Wikipedia Articles | 229 | |
| Conc | clusions | | .231 |
| Intro | duction | | .232 |
| | | | |
| Cha _j | pter 7 Conclusions Thesis' Objectives | | 232 |
| | 1.1 The validity of modelling Trust is a form of defeasible reasoning | | . 232 |
| | 1.2 Designing a trust model based on Defeasible Reasoning | | |
| | 1.3 The positive impact: what to expect and how to evaluate it | | |
| 7.2 | Contributions | | 238 |
| 7.2 | Discussion of the method and open issues | | |
| 7.3 | | | . 270 |
| | Classes of Applications and Ongoing Projects | | |
| | 3.2 Similarities with other computer science applications <i>modus operandi</i> | | |
| / | 5.2 Similarities with other computer science applications modus operanal | 242 | |

| 7.3.3 Features of the computation | 243 |
|--|-----|
| Application-Contained nature of the computation | 243 |
| Justifications | 244 |
| Argumentation Scenario | 244 |
| Alternative to the couple past-outcomes/Recommendation | 244 |
| Evidence Selection | 245 |
| 7.3.4 Open Issues and Future Development | 245 |
| The phase of Domain Analysis | 245 |
| Mapping, Critical Question and Computation | 246 |
| Conclusions | 250 |
| | |
| Relevant Author's publications | 251 |
| References and Bibliography | 251 |
| | |
| Appendices Appendix A Computational Models of Trust | 261 |
| Computing trust using Past Outcome and Direct Experience | |
| The Theory of Probability as a Computational Tool in Trust | |
| Computing trust using Indirect Experience | |
| Game Theory: the utility game | |
| Using Risk to compute Trust | |
| Cognitive models, the Falcone-Castelfranchi model | |
| Trust-Based Heuristics: two examples | |
| Monitoring the system: the model by Carter | |
| Miscellanea | |
| Appendix B Statistical Tools Used | |
| Basic Statistical Operators | |
| Statistical Tests and Concepts | |
| Statistical significance tests | |
| Sampling | |
| Correlation/Relation among variable | |
| Order and Ranking Method | |
| Confounding Variables | |
| Appendix C Trust Schemes and Critical Questions: a detailed discussion | |
| How trust schemes are described | 285 |

| Time-based Trust Schemes | 286 | |
|--|-----|--|
| Longevity | 286 | |
| Trust Scheme Persistency/Regularity | 287 | |
| Trust Scheme Stability | 291 | |
| Trust Scheme based on Information Sharing and Social Role | 292 | |
| Information Sharing | 292 | |
| Trust Scheme Recommendation or Indirect Experience | 293 | |
| Trust Scheme Reputation | 295 | |
| Trust Schemes linked to Social Role | 296 | |
| Trust Scheme Authority | 296 | |
| Trust Scheme Connectivity | 298 | |
| Trust Scheme Popularity | 299 | |
| Trust Scheme visibility/accessibility | 299 | |
| Trust Scheme Transitivity | 299 | |
| Trust Scheme Information Provision. | 300 | |
| Trust Schemes linked to Activity | 301 | |
| Trust Scheme Pluralism | 301 | |
| Trust scheme Activity (trust based on the degree of activity of an entity) | 303 | |
| Trust Scheme linked to outcomes | | |
| Trust Scheme based on prejudices and grouping | | |
| Similarity, Categorization, Standard | | |
| Similarity to Trust | 310 | |
| Game theoretical/Cognitive Trust Scheme | 311 | |
| Trust Scheme common goal/situation/risk | | |
| Benefits/Costs Trust Scheme: the trustier motivation | | |
| Trust Scheme fulfilment | 312 | |
| Trust scheme derived from exogenous factor: risk and disposition | 313 | |
| Trust Scheme Risk | 313 | |
| Appendix D Plausbility, amount of knowledge and aents ranking | 314 | |
| Appendix E Statistical method to isolate confounding variables | | |
| Appendix F List of Trading Terms used in Competence Analysis | | |
| Appendix G Principal Component Analysis for Experimet IV | 329 | |

List of Figures and Tables

| Chap | ter 1 | |
|------|--|---|
| | Figure 1.1 Figure 1.2 Figure 1.3 | The basic scenario: from evidence to a trust-based decision Trust Model design's issues Hypotheses to be evaluated. |
| | Figure 1.4 Figure 1.5 | High-level architecture of our model Evaluation strategy Flow-Diagram |
| Chap | ter 2 | |
| | Figure 2.1 | Trust value Probability Distribution |
| | Figure 2.2 | Subjective Logic Triangle |
| | Figure 2.3 | A computational Trust Solution |
| | Table 2.1 | Uncertainty in value conversion |
| Chap | ter 3 | |
| | Figure 3.1 | Presumptive reasoning to sustain the plausibility of a topic in a specific context. |
| | Figure 3.2 | Inference graph for rules 6 and 7 |
| | Figure 3.3 | D is a rebuttal defeaters while C is an Undercutting Defeaters |
| | Figure 3.4 | An example of inference graph (left). Q and ¬ cannot be recursively computed (right) |
| | Figure 3.5 | A more problematic situation, and the effect of an external argument T |
| | Figure 3.6 | Quantifying the effect of defeaters |
| | Figure 3.7 | Circular Path |
| Chap | ter 4 | |
| | Figure 4.1 | Analogy between Walton presumptive reasoning and Trust Evidence selection |
| | Figure 4.2 | Trust reasoning: from facts to trust |
| | Figure 4.3 | Conclusions derived from premise P with reason-link R1 |
| | Figure 4.4 | Conclusions derived from two separate reasons |
| | Figure 4.5 | Conclusion C derived from the premise (P1 and P2) |
| | Figure 4.6 | Supporters and defeaters |
| | Figure 4.7 | A deductive trust scheme |
| | Figure 4.8 | Trust scheme past-outcomes in a given epistemological state |
| | Figure 4.9 | A model of Trust based on Defeasible and Presumptive Reasoning |
| | Figure 4.10 | A model of Trust based on Defeasible and Presumptive Reasoning /2 |
| | Figure 1 11 | Evidence collected about Mark's exams |

| Figure 4.12 | A defeasible reasoning link |
|-------------|---|
| Figure 4.13 | Supporter and Defeater |
| Figure 4.14 | Two independent arguments supporting the same conclusion |
| Figure 4.15 | Defeaters and Supporters |
| Figure 4.16 | Circular Path |
| | |
| Chapter 5 | |
| Figure 5.1 | Rank-based strategy |
| Figure 5.2 | Mutual relationships among arguments |
| Figure 5.3 | Function of conclusion, attack, support |
| 118010 3.3 | Tanetton of conclusion, attack, support |
| Table 5.1 | Time-based Trust Schemes |
| Table 5.1 | Information-Sharing Trust Schemes |
| Table 5.2 | Social-Role based Trust Scheme |
| Table 5.4 | Activity-based Trust Scheme |
| Table 5.5 | Outcome-based Trust Scheme |
| Table 5.6 | Prejudges- and Grouping-based Trust Schemes |
| Table 5.7 | Game Theoretical-based Trust Schemes |
| Table 5.8 | Summary of Trust Schemes |
| Table 5.9 | Mutual relationships among arguments |
| Table 5.10 | Function of Support analysis |
| Table 5.11 | Function of Attack analysis |
| 1 doie 3.11 | Tunetion of Attack unarysis |
| Chapter 6 | |
| Figure 6.1 | Evaluation strategy |
| Figure 6.2 | FinanzaOnline.it application model |
| Figure 6.3 | Some visible data of a forum member |
| Figure 6.4 | Percentage of Members according to their Reputation Value |
| Figure 6.5 | Percentage of Members according to their Reputation Value (active |
| | members only) |
| Figure 6.6 | Wikipedia Application Model |
| Figure 6.7 | Featured and Standard Articles distribution |
| Table 6.1 | List of Experiments |
| Table 6.2 | Pool results |
| Table 6.3 | Matching data and trust scheme |
| Table 6.4 | Persistency Trust Scheme analysis |
| Table 6.5 | Activity Trust Scheme analysis |
| Table 6.6 | Competence Trust Scheme analysis |
| Table 6.7 | Connectivity Trust Scheme analysis |
| Table 6.8 | Experiment I overall results |
| Table 6.9 | Experiment I overall results (2) |
| Table 6.10 | Experiment II overall results (1) |
| Table 6.11 | Experiment II: Argumentation vs. Aggregation |

| Table 6.12 Table 6.13 Table 6.14 Table 6.15 Table 6.16 Table 6.17 Table 6.18 Table 6.19 | Experiment III: Messages collected Experiment III: Percentage of Good suggestion Experiment III: Good suggestions divided by Risk level Experiment IV: Argumentation vs. Aggregation Experiment V overall results Experiment V overall results (2) Experiment VI overall results Experiment VII overall results |
|---|---|
| Chapter 7 | |
| Figure 7.1 Figure 7.2 | Trust Model design's issues Hypotheses to be evaluated. |
| Appendices | |
| Figure A.1 Figure A.2 Figure A.3 Figure A.4 | Generic shape of the update loop of Direct Experience Beta Distribution family. Plausibility and Belief in A based on evidence E _i High Level Architecture of the SECURE trust engine Initial Trust value assignment in Sierra. |
| Figure B.1 | The function R(n) Langavity Informed Graph at a certain enjotemalogical state |
| Figure C.1 Figure C.2 | Longevity Inference Graph at a certain epistemological state Activity Inference Graph. |

Chapter 1 Introduction

1.1 Trust, from Social to Computer Science

Today's digital world is distributed, collaborative and open. New emerging systems such as wikis, blogs, marketplaces are populated by entities previously unknown to each other. Examples include e-business applications, online auctions, online web communities, wiki-based and other collaborative environments, wireless open network, and peer-to-peer system. In these environments cooperation with unknown entities is likely, representing both a potential danger and an opportunity - seldom an obliged choice. These interactions may involve lack of control or dependency of one of the two parties in favour of the other, which is therefore in a position where it can affect the other.

Computational models of Trust seek to exploit the human notion of trust in open and collaborative environments. The basic assumption is the following: since trust is undoubtedly a successful fundamental element of human relationships, similar benefits are expected by embedding computational model of trust into digital worlds.

The community of researchers accepts that trust has a distinctive and recognizable nature, but, as any human phenomenon, it is influenced by multiple factors such as culture, history, sociocognitive aspects and so forth. The use of the word trust in computer science reflects this fuzziness. The term is used in *trusted computing* referring to computer systems that act as

expected and behave consistently with a set of specifications. Since this behaviour is enforced and protected with security mechanisms, the meaning of the world trusted in this context contrasts with the common-sense notion of trust. The term *safe* computing could have been more appropriate. As various authors noticed [Gam00], trust has meaning only in situation where there is a lack of control and the trustee entity can freely choice between betray or act honestly,

The term *Trust* is also used in applications such as private-key infrastructure. In these scenarios, users accept to trust digital certificate signed by trusted authority. Despite the reductive and too simplified notion of trust underlying these systems, the term trust is closer to its human notion.

Trust systems encompass also successful application such as online reputation systems. These systems are genuine based on the notion of reputation, a complex ingredient of human trust.

Anyway, it is good to underline how these systems manipulate and users' explicit feedbacks, and therefore they are better described as tools that allow elaborating and propagating trust, rather than tools to *produce* or *reason extensively* about trust. In the basic idea of Reputation Systems, *Trust* is hidden and delegated in users' mind.

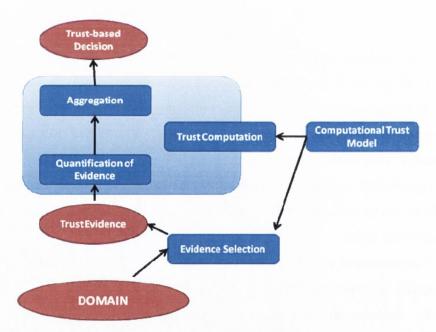


Figure 1.1 - The basic scenario: from evidence to a trust-based decision

This thesis intends the world *Computational Trust* as the study of how the human notion of trust can be modelled and exploited to support agents' decision making process. The computation generates trust values, reasons able to help users/agents in the decision making

process in open and collaborative environments. The scenario, depicted in figure I.1, is a classical decision-support problem. An agent – the *trustier* - is acting in a certain domain where he needs to trust other agents, whose ability and reliability are unknown. Agent collects the available *trust evidence* from the domain and it quantifies it by mean of a *trust computation* informed by an underlying *model of trust*. The computation involves also the critical stage of *aggregating* various evidence in order to obtain a final trust value to be used in its decision making process. If there are enough reasons, trust is granted and the interaction starts. Therefore we intend trust as a support for decision making: a good trust model is the one that produces values that help agents making correct decisions.

Despite the proposed scenario is the base-line situation for any trust-based decision support tools, in many situations it encompasses complex issues that are still open challenges in trust computations. Challenges are in the selection of relevant elements representing evidence for trust and its justification, the analysis and quantification of the evidence, how to aggregate multiple evidence into a final trust suggestion. These three main issues have to be clarified, since they also define the problem space of this thesis.

Evidence Selection

Evidence selection is the problem of deciding which elements may have a meaning for trust and therefore be used in a trust computation. We identified two main problems from the state-or-the-art review described in chapter II and IV. The first is "the need for a clear process between trust models and trust evidences" [Sei06], that we could rewrite as the possibility of defining a general methodology to accomplish evidence selection in trust. The second problem is the limited number of types and source of trust evidence that are considered in computational trust. For instance, let's consider the trust systems mentioned above. Recommendations-based systems avoid the problem of evidence selection, since this complex task is delegated to users providing feedbacks. Systems such as PKXI [Bal03] have a dedicated trust infrastructure where evidence is well-defined objects known a priori, such as a digital certificate, and therefore the problem of evidence selection is rather a problem of evidence gathering. When evidence is directly identified and assembled over domain elements, evidence selection is a challenge: a justification for the selection is required and the process may result asystematic, unbounded and subjective.

Trust Computation and mechanisms

The second issue regards the way trust is computed and the notion of trust encoded in the computational mechanisms.

We contend that there is a gap between rich theoretical or cognitive models and the current landscape of computational models of trust, and the consequent limited role of a *generic expertise* of trust in the process.

Rich of Cognitive formal models of trust have been introduced from the very beginning of computational trust studies, as Marsh's first model is an example. Nevertheless, these models are often criticized for being a high-level view of trust that does not have feasible computational implementations, appearing as a pointer to computations that cannot be performed and evaluated. On the other hand, computational systems have preferred to follow a reductionist and *hard approach* to trust. In these approaches, the role of trust as a human phenomenon is narrow; the collection of evidence is delegated to users, encoded in a dedicated infrastructure, simplified or delegated to experts of the domain rather than *experts of trust*. Trust computation is often a rigid formula approach, containing limited trust intelligence. In line with authors such as Marsh [Mar94] or Castelfranchi [Cas00] we reject any reductionist approach, and we consider trust as a complex phenomenon that calls for a soft approach able to produce human-understandable justifications and reason about arguments or exceptions, a process that is context-aware computable.

Aggregation of Evidence

The hard approach to trust is evident in the way multiple pieces of evidence are aggregated in a final trust decision. A trust computation usually collects several pieces of evidence that have to be aggregated into a final value. The present state of this computation is usually performed with an averaging approach that, even present in complicated variants, has the common assumptions that evidence is independent from one other, isolated objects that do not interfere, and therefore it can be aggregated by a simple aggregation formula. This implies an approach where contradictions are blended, and the obtained values are slow-to-react average.

On the contrary, we content that the understanding of the mutual relationships among evidence in trust is essential in order to perform a correct decision, and it reflects the way humans think, where contradictions are important source of information and decisions can be rapidly reverted.

The model described in this thesis is aware of and it confronts itself with the above three

challenges. Our main goal is to capture the way humans decide to trust and translate it into a computation. One of our assumptions is therefore the rejection of any reductionism approach, and the idea of trust as a complex form of reasoning performed over a network of evidence.

Trust is intended as a form of reasoning performed over plausible evidence. The arguments used in this reasoning are explicit, forming a recognizable expertise of trust that supports evidence selection and quantifications. These arguments are intended as defeasible mechanisms, meaning that are not valid per se but assumptions whose plausibility has to be studied. Finally, the model does not consider the different evidence in isolation, but it considers how one piece of evidence may support and attack other pieces, resulting in a more human-like method of reasoning that producing more justifiable decisions.

The next section starts introducing the core issue of our thesis, presenting the starting idea of introducing the theory of defeasible reasoning in trust.

1.2 Our Starting Point: trust as a defeasible phenomenon

In order to introduce our research issue, we first need to describe the background idea underlying it. Our starting point is the simple observation that the way humans grant, deny or evaluate trust is defeasible. The term *defeasible* comes from *Argumentation Theory*, the multidisciplinary study of humans' way of reasoning and discussing.

Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid [Sta05]. The truth of the premises of a good defeasible argument provides support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information.

Let's consider this assertion:

If a number can be divided by four, therefore it can be divided by two

This proposition is deductively valid. If the premises are true, the conclusion follows necessarily and even if we add extra premises, the conclusions do not change. For instance, adding that the number is greater than ten or the number is negative does not change the conclusion, which is fully included in the truth of the premise "the number can be divided by four".

Let's now consider the proposition:

The grass was wet this morning, so it rained overnight

We agree that even if the premise is true and the grass is actually wet, the conclusion could be false. The conclusion seems reasonable if we do not know anything else, but suppose that we noticed that the road was wet, we could argue that maybe there is another explanation for the wet grass, maybe the sprinkler was on [Bre97]. We have just reasoned in a defeasible way: by adding extra information, the link between premise and conclusion changes – this time it becomes weaker. Formally, given a true proposition *A*, deductively valid inference follows this rule:

if
$$A \vdash p$$
 then $A, B \vdash p$

Therefore, even adding extra premises, the conclusion p stands. A defeasible proposition does not follow the above property, hence the name of non-monotonic given to this kind of reasoning, since new premises may change the conclusion p.

Going back to trust, let's consider the following two propositions used to sustain a trust decision:

- a) I assigned this computation to Mark, he passed 8 of his 10 last maths test
- b) I suggest you the baker's shop over there, in 20 years I have never seen it without customers

In the first proposition there is the concept of delegation, in the second a recommendation. They are both common situations in trust.

In the proposition a trust is justified by relying on past-outcomes of an entity: Mark should be good in performing the requiring computation since it passed 80% of his Maths tests, while in proposition b the butcher's shop is suggested because of its long and constant success with its customers.

Which kind of propositions are *a* and *b*?

They are clearly *defeasible* propositions, even if in trust models – especially the first one – they are never considered so.

For instance, what if we know that Mark failed only two tests, but they were the last two he did, while the first tests were generally easier? 80% of tests passed are now a weaker evidence.

Regarding the butcher's shop, what if we know that the shop is the only one in the range of 20 km? This lack of competitors makes the high number of customer a weaker argument for trust.

Today landscape of computational models of trust considers the mechanisms they employ deductively valid. The use of past-outcomes mechanisms or recommendations covers the majority of the systems, in which no procedure to test the plausibility of the conclusions is provided, and a hard approach to the computation is kept. Few mechanisms encompass

uncertainty management, but uncertainty is a separate concept from defeasibility. We could be perfectly certain of the data used for taking a trust-based decision, but still the reason employed can be highly implausible and defeasible.

As we describe in our state-of-the-art review, only some cognitive-inspired models are aware of this defeasible nature, even if not explicitly stated and no implementation are known to the author.

A second problem is the way different pieces of evidence used in trust are aggregated. In general, a trustee may have more than one reason on which a decision should be taken. Evidence can be contrasting, some in favour and some against the trustee.

For instance, there could be 6 evidence in favour of John and 2 against. For simplicity let's suppose that all the evidence has the same strength, all set to +I (positive) and -I (negative).

The simplest approach is to sum the evidence producing a final trust value of 6-2=+4, or the percentage of positive evidence, equal to 6/8=75%.

We refer to this procedure as *simple aggregation*. There is one feature that defines this strategy: different evidence is considered separated from each other and kept isolated. The mutual relationship that may occur between them is not considered. In the best situation, more importance can be assigned to a piece of evidence than another. This is a first step in the direction of recognizing that evidence is more plausible than another, but still the process is seldom justified or left to intuition, and again the mutual relationship of the evidence is not considered.

Considering the mutual relationship among evidence means to investigate if the presence of evidence A and its value (positive or negative) influence, for good or for bad, another evidence B. It might be the case that that one of the two negative pieces of evidence about John invalidates 3 of the positive, so that these 3 are nullified. Therefore, the actual evidence to be used are 5, 2 negative and 3 positive, with a resulting trust value of +1 or 3/5=60%, strongly different from the previous one.

Suppose that Alice has to trust John in a context in which *longevity* – the fact that an entity is in the specific environment for a long time – and the *degree of activity* – how many interactions an entity accomplished in his lifetime – are both plausible evidence for supporting trust, as Alice's investigation has proven.

Alice collects the following contrasting evidence: John has high *longevity* but a low *degree of activity*. Which is Alice's final decision? Alice could ignore any link between the two pieces of evidence, and conclude that I-I=0, and she has no reason to trust or distrust John. Anyway, by

looking at the evidence, the high degree of longevity of John loses meaning when we consider its lack of activity: even if John has been in the environment for a long time, he does not interact as he was supposed to do. A lower degree of longevity but with higher activity could have been a better situation. The evidence *activity* defeats the evidence *longevity* and Alice has now a good reason to distrust John.

This procedure takes advantage of the mutual relationship among evidence that requires gathering extra knowledge about the nature of the evidence.

Of course it might be the case that such knowledge may be not available or accessible. Therefore, a *simple aggregation* strategy is justified in conditions of ignorance about the mutual relationship of the evidence, or when it is known that all pieces of evidence are completely independent.

The two examples – the defeasibility of trust mechanisms and evidence aggregation – are aspects of the same procedure, a defeasible argumentation. This division has been made explicit because of the way we adapted the selected theory in our trust model, but they are actually two stages of a single process. In the first one we investigate the plausibility of a specific piece of evidence, in the second the plausibility of the evidence in presence of other evidence. The common core issue is that an evidence or assumption A should be analysed by taking account of how another evidence E can support or defeat it, the set of evidence E being internal conditions related to the reasons supporting E, or separated evidence that influences E conclusions.

1.3 Research Issue

After having introduced the main rationale behind computational trust and the concepts of defeasible reasoning, in this section we formulate and comment the research issue investigated in this thesis. The basic definition of our research issue is the following:

Since trust is a form of defeasible reasoning, presumptive reasoning and defeasible reasoning techniques have a positive impact on the quantitative analysis of trust.

If trust is a form of defeasible reasoning, the computational techniques that researchers in AI and Argumentation Theory have developed for this kind of reasoning could be effective in trust computation.

If trust is a defeasible phenomenon, the mechanisms underlying a trust-based decision should be treated and computed as a defeasible argument that can be defeated or supported according to the current epistemological state. Moreover, the multiple evidence composing a trust-based decision is in general mutually influenced, and therefore its aggregation should consider this mutual relationships.

The new defeasible-based computation should increase the quality of predictions based on computed trust computations.

Our starting research issue is now analysed and better defined. The research hypothesis implies several issues:

- 1. The validity of the starting assumption trust is a form of defeasible reasoning
- 2. How to use defeasible reasoning and presumptive reasoning techniques in trust, i.e. how to define/design a computable model of trust structured over these two disciplines.
- 3. To define the positive (if any) impacts of the envisaged model: what we expect, why and how results derive from the introduction of Defeasible Reasoning, to define how to quantify and evaluate the expected positive impacts.

The above issues contain both a theoretical contribution to the definition of a model of trust and a practical contribution in evaluating the effect that such a model has over trust computation: does it make trust computation more efficient? Does it introduce novel computational paradigm? Let's continue the incremental definition of our research issue by considering each of the above three points.

Issue 1: The starting assumption: trust is a form of defeasible reasoning

Instead of taking our starting assumption for granted, we address the validity of it.

We need to investigate and show how our starting assumption is reasonable and therefore can be accepted. It contains actually two steps: (i) trust is a form of reasoning, and (ii) this reasoning is defeasible. In chapter III we provide arguments for both this issues. The large majority of social scientist definitions of trust agrees that trust may have the form of a complex evaluation or a *reasoning*. This means that trust has also other forms – unconscious decision, habit, intuition – but reasoning is an acceptable form. In particular, in the context of rational agents computing trust, representing trust as a form of reasoning seems the most adequate approach. Representing trust as a form of reasoning is a first step towards ruling out the reductionism approach that many computational models suffer of.

The other issue, trust is a defeasible form of reasoning, is also investigated with arguments similar to the ones presented in the previous section.

Issue 2: How to Design a Model of Trust as a form of defeasible reasoning

The issue calls for a model of trust able to take advantage of defeasible reasoning theory and techniques. It is required to identify the adequate techniques from the defeasible reasoning literature and to define how they can be used in trust computation. We anticipate the two techniques used: Walton's presumptive reasoning and Pollock's semantic for defeasible reasoning. The first theory is a more descriptive approach; the second is a formal semantic. Both of them are needed to model trust.

The problem of adapting these two theories to trust is not trivial and it produces our main theoretical contribution. In order to compute trust as a form of defeasible reasoning we need to identify which are the generic arguments used in a trust-based decision. This requires a detailed investigation of the current landscape of computational models of trust in order to identify the recurrent patterns—if any- underlying computational trust mechanisms. Moreover, it requires the investigation of state-of-the-art of research in trust and cognitive models in order to identify new mechanisms accepted by researchers but not yet computationally investigated. In other words, we state that not only is trust a reasoning, but it is a reasoning composed by recurrent patterns, a kind of ingredients of trust. This assertion is again well grounded in social science and particularly in the Cognitive model, where trust is made of declared components.

The identified mechanisms are not yet ready to be used in a defeasible reasoning. We need to produce a version of them as *defeasible* arguments. This implies investigating assumptions underlying each mechanism, identifying possible supports or defeaters arguments, and a way to assess their plausibility.

These mechanisms – existing or new – in their defeasible versions form the arguments of the trust-based defeasible reasoning.

Finally, the last issue is to identify, among the identified trust arguments, mutual relationships and define a formal semantic that describes how to combine contrasting and supporting arguments. The list of issues is represented in figure 1.3.

Chapter III contains the overall design of our solution, chapter IV our state-of-the-art review of computational trust mechanisms and chapter V our list of defeasible trust arguments.

We stress a key concept of this thesis: as we reject any reductionism approach, we believe that trust is an expertise per se and it is therefore possible to define its mechanisms. We could have left this assumption outside this thesis, and provided just a framework for computing trust based

on defeasible reasoning leaving the definition of the arguments unknown. This thesis's contribution would have been largely reduced. Had we avoided defining generic arguments to be used in trust-based reasoning, generic techniques of defeasible reasoning would have sufficed.

On the contrary, our research issue is to model that specific form of defeasible reasoning that is trust, with its recurrent patterns, mechanisms and semantic.

Moreover, the explicit definition of trust arguments is justified and required by Argumentation Theory and by Walton's theory that we want to adapt to trust. A basic and largely accepted assumption of these theories is the existence of recurrent pattern of reasoning called argumentation scheme. Therefore, it is one of our main tasks and research issues to identify those generic argumentation schemes used in a trust-based decision that we call trust schemes.

On the other hand, we do not claim that we provide a definitive and comprehensive list of trust mechanisms, but we claim that generic trust mechanisms exist, we provide a detailed list and we prove their efficiency in the computation. Note also that, in accordance with the main idea of this thesis, we reason in a defeasible way: each mechanism identified is valid but subject to defeat, and therefore not an absolute statement.

Issue 3: The positive impact; what to expect and how to evaluate it

The generic term positive impact has now to be defined in details. With positive impact we refer to the efficiency of the trust value computed. Since trust value is a prediction about the trustworthiness of an entity, a trust metric A - a collection of entities and their computed trust values - is more *efficient* (or *valid*) than a *trust metric B* when the number of *true positive* and *false negative* generated by A is greater than B and the number of *true negative* and *false positive* is smaller.

Trust metric A can be proven to be efficient also if its predictions are similar to another trust metric C that is known and accepted to be of high value. The two trust metrics have to be produced in a complete independent way. This second procedure is the strategy we use in our evaluation: we prove that our model's computation is efficient by comparing it with an external and independent trust metric recognized to be valid. More details of our evaluation strategy are given in section V of this introductive chapter.

Now we focus on which positive impacts are expected by the introduction of defeasible reasoning.

First, we expect that the introduction of our defeasibility study of computational mechanisms used – as designed in our model using Walton's presumptive reasoning- will positively impact the trust computation by giving more importance to plausible mechanisms and by invalidating implausible ones. This increases the true positive predictions –since plausible arguments count more - and reduces the true negative ones – since implausible arguments are defeated. In other words, we need to show that a trust metric whose mechanisms are subject to a defeasible analysis of its plausibility is more effective than the same trust metric considered deductively valid. Due to its defeasible nature, we need to show how its efficiency changes according to its plausibility. This hypothesis is also extensible to a single trust mechanism. We refer to this as **hypothesis 1**: defeasibility analysis makes trust mechanisms more efficient.

We expect that the application of our defeasible reasoning semantic for trust, modelled after Pollock's generic one, will impact the way multiple evidence is aggregated. More precisely, we expect that a final trust metric based on a defeasible reasoning semantic is more efficient than a metric generated by a simple aggregation strategy .We refer to this as **hypothesis 2**.

Finally, we made explicit also the following **hypothesis 3:** the overall method has to provide efficient trust values. Not only does defeasible reasoning have to increase the efficiency of the trust value (hypothesis 1 and 2) but we have to show our defeasible computation is efficient.

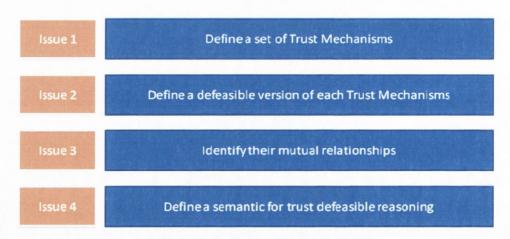


Figure 1.2 - Trust Model design's issues

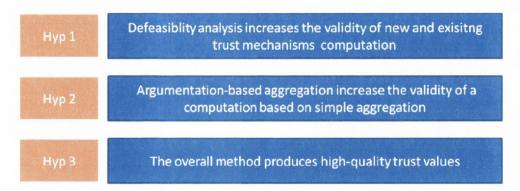


Figure 1.3 - Hypotheses to be evaluated.

1.4 Model Quick Overview

In this section we briefly introduce the key ideas behind the design of our model, depicted in figure 1.4. The model intends trust as reasoning over plausible evidence. Evidence is collected in a pattern-matching fashion between available domain elements coming from the application with a list of generic *trust schemes*, our defeasible valid reasons to trust an entity. For instance, a trust scheme is the past-outcome one: "X trust Y because previous interactions between X and Y were positive".

Our list of trust schemes encompasses outcome-based schemes, information sharing ones, social-based, temporal factors such as accountability and persistency, activity-based and so forth.

As a result of this matching, we have a list of *trust evidence* whose plausibility has to be investigated. In our model, each trust scheme T_s has a set of critical questions attached to it, expressing arguments that can defeat or support the trust scheme and the corresponding evidence identified by the scheme. This operation may require access to further domain elements.

After the plausibility study, the trust arguments, in favour or against the trustee, enter a defeasible reasoning computation, simply called the argumentation. Here the trust arguments are aggregated by considering their mutual relationship to generate a set of sustainable conclusions.

The final stage of the process is the trust-based decision, function of:

- The output of the trust-based reasoning
- The trustier' risk assessment of the specific situation and its disposition to risk
- The trustier's disposition to trust

The model is depicted in the figure below. We call matched trust schemes domain elements that have been matched by one of the trust scheme, and trust arguments the schemes after the

defeasibility analysis. Note that not all the trust schemes enter the argumentation phase, since some of them are discarded because considered implausible, and not all the trust arguments are used in the final trust decisions, since some of them may have been invalidated by other trust arguments during the argumentation.

Our trust expertise is clearly marked and composed by a set of trust schemes and their critical questions.

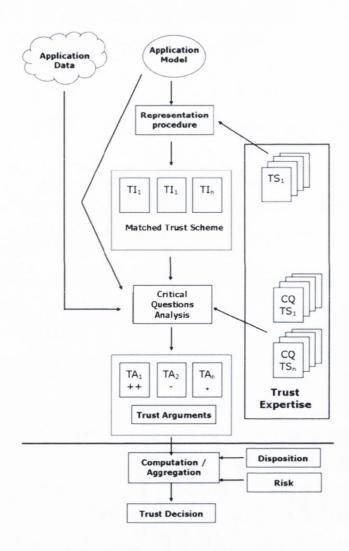


Figure 1.4 - High-level architecture of our model

1.5 Evaluation Strategy: comparative evaluation of Trust Metrics

Evaluating trust reflects the complex and fuzzy nature of this concept. The consequence is

that a clear evaluation strategy in trust does not exist. The diverse nature of trust models makes their evaluation goals different: evaluating an online reputation system differs from evaluating a trust algorithm in the context of security: the first focuses on the utility of recommendation, the second on the robustness of the algorithm if attacked.

Since it is the human notion of trust that our model wants to capture and quantify, an evaluation with real data is preferred to simulation-based approach. Our problem is to answer the following question: how can a computational model of the human notion of trust be evaluated? When is a model considered of good quality? How can we compare two models?

The answers start from the simple assumption that a computational model of trust makes predictions about the trustworthiness of entities. A good model of trust is therefore the one able to make correct predictions, supporting a decision-making process by assigning good value to trustworthy entities and bad value to untrustworthy ones.

The issue is therefore what has to be predicted.

When humans grant trust, they expect that the trustee will act in a way that produces a desirable outcome that fulfils their needs [Gam00]. Therefore, what has to be predicted is that ability to produce expected outcomes, behave as expected or deliver certain results.

There is an underlying assumption in the above proposition: there is a common understanding between the *trustee* and the *trustier* about what has to be considered *good* and *desirable* outcomes/behaviour.

If we accept that trust is a purely subjective phenomenon, it does not make sense evaluating trust models, since good or bad predictions are valid only in the context of a single *trustier* entity. If trust is purely subjective, it cannot be transferred, and concepts such as reputation or trust transitivity lose meaning.

We believe that trust has at least a *weak* objectivity. Many authors, notably Gambetta [Gam00], showed how trust exists as a fundamental glue of human society and therefore it is linked, reflect and promote common values of a society. This objective nature of trust is what makes an evaluation possible.

Therefore, evaluating trust makes sense in the context of a community if there is a plausible *common understanding* about roles, behaviour and values. Note how this implies that trust cannot be evaluated in a multi-agent society when this common understanding is uncertain or weak. It does not make sense to evaluate a reputation system about a highly subjective matter.

Our first conclusion is that any trust evaluation has to start with an investigation of the existence

of a plausible *common understanding*.

When a common understanding can be postulated, a trust metric can be evaluated by quantifying how well it is able to predict it. We identify three ways for achieving this result: (i) the common understanding can be already encoded in an existing trust system, or (ii) it can be defined with the explicit help of the community or (iii) it can be made explicit by defining a set of outcomes and behaviours compatible with the common understanding.

In the first case, we actually have available an existing trust metric. If an existing trust system already in place in the context, that must be well accepted by the community, useful and tested. If such a system exists, a *comparative evaluation* can be performed by quantifying how much the two trust metrics are similar. We used this strategy in the experiments performed in the context of the Wikipedia project, relying on an internal mechanism of Wikipedia that assigns to its articles and authors quality awards.

The second possibility is to build a trust metric by asking direct help to the community members. For instance, a trust metric can be defined with a poll over the community members, collecting a trust-ranking of entities populating the environment.

Then, a *comparative evaluation* is performed as in the first case. Of course, the trust metric build upon direct feedback coming from users community must plausible, accepted by the community, based on a reasonable numbers of votes and voters.

There is an essential hypothesis to be respected when performing a comparative evaluation: the trust metric used for comparison must be generated independently and the mechanisms used do not have to overlap. For instance, if an existing trust metric has been defined using explicit opinions from community members, a recommendation system cannot be used as an example of independent metric, since the mechanism underlying the computations are clearly linked to members 'feedback.

In the third case, having identified a plausible common understanding, a model can be tested not with a comparative evaluation with another trust metric, but rather with a direct comparison with outcomes of the interaction.

The strategy implies the following steps:

- defining a set of expected outcomes feasible under the hypothesis of common understanding –computing our model's trust metric,
- performing predictions regarding interactions using the values computed by the metric,
- measuring the rate of correct predictions by collecting the outcomes of the interactions.

Therefore, a trust metric is now defined with on an absolute scale, as the percentage of good predictions.

For instance, in the context of an online trading community, where members share news, analysis and suggestions about trading, it is reasonable that, if member A suggest to buy company X, and X increases its price after the suggestion, that is definitely a good outcome. The domain is certainly one where a common understanding can be postulated, and the common goal is clear.

Note that this evaluation is not more objective that the previous. Relying on outcomes could seem a strong and straightforward method, but it could be reductive and not completely correct, as many author pointed out [Cas02] and we also argue in chapter V. It could be argued that the above example capture only one aspect of trust in that context, aspect certainly agree by almost all the members but not the only reason to trust. It is again a defeasible argument that, nevertheless, could have a strong validity in several domains, where obtaining positive outcomes is regarded by the community members a necessary condition for trust, as for the experiments we perform in the context of *FinanzaOnline.it*. Figure 1.5 defines our evaluation flow diagram.

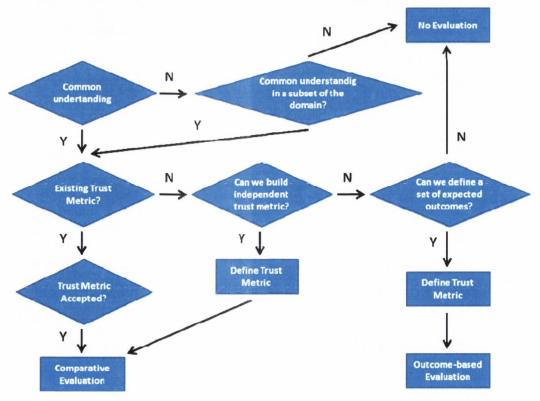


Figure 1.5 - Evaluation strategy Flow-Diagram

1.6 Thesis Contributions

This section describes the contributions of our work. We divided contributions into *theoretical* contributions, deriving from our method's design, and in *computational* ones, related to the effect of the model over the way trust is computed and the output obtained from our computations.

The main results could be summarized as follows: by starting from a set of generic reason to trust, by mapping these patterns to application elements, and by reasoning defeasibly about their validity and by combining them considering their logical consistency, we have obtained trust values of very high precision.

Computational contributions

- Introduction of defeasibility (main)
 - We introduced defeasible computational techniques and we adapted them to trust. The resulting computation has produced:
 - o A more efficient defeasible aggregation strategy than the previous one
 - A more efficient use of existing trust mechanisms due to the introduction of their plausibility study, a direct consequence of taking the mechanisms as defeasible arguments
- We introduced new mechanisms to compute trust, namely temporal factors, activitybased and stability-based mechanisms, as better described in chapter IV and VI, that we proved to be efficient
- An overall model that produces trust value of very high quality, values able to predict with what humans consider trustworthy, as proven by our experiments
- A trust computation behind the usual two mechanisms, recommendations and pastoutcomes. The two mechanisms, lauded are objective and covering almost 95% of
 computational trust applications (see chapter VIII), were critically analysed showing their
 defeasible nature and potential inefficiency. Our computation provides alternative metric
 to be used to check the validity of these two schemes, or extra data to be merged in a
 more efficient decision-making process.
- A tool for computing a trust-based data-mining. As we shown in the evaluation section, the model suits applications where a large set of data representing behaviour of a community needs to be mined for extracting trust-related information
- An example of application-contained computation, where trust is computed by relying on

elements inherent to the application, without relying on an explicit trust infrastructure or humans' judgements. The computations have two main points: not invasive and it can be used as a valid comparison with alternative mechanisms such as recommendations-based systems

• Evaluation in real scenarios. This thesis describes the first extensive trust-based computation on the context of the Wikipedia project.

Theoretical

- Trust schemes and Critical questions. We defined an extensive list of trust schemes. The majority of the schemes are existing trust mechanisms extracted from literature, some represent a computational novelty. Our contribution for the first set of schemes is the definition of their defeasible version, i.e. the study of the assumptions they rely on. For the second set of trust scheme, we introduce a computational version of them.
- A semantic for trust reasoning. We contributed both trust model definitions and to Argumentation Theory by defining a defeasible reasoning semantic for trust, which specific structures and computation
- Evidence selection. Our model is a step in the definition of a general methodology for evidence selection, and a consequent investigation of a new set and source of evidence. By starting from a representation of the domain, our model identifies elements that could be trust evidence. Our selection is driven by our trust scheme, a generic set of identified mechanisms that humans adopt to sustain a trust decision. The key issue is that our trust model contains a generic trust expertise that justifies the evidence selection. Instead of relying entirely on domain-specific expertise, we assumed that trust is an expertise per se that by its own is able to sustain a valid evidence selection. Domain-specific expertise has a supportive role instead of delivering the solution. Moreover, the generic trust expertise is able to sustain a methodology, giving that systematicity that subjective ad-hoc heuristics approach defects.
- Argumentation. Our model introduces the idea of trust as an argumentation between two
 parties, the trustier and the trustee. This dynamics could be implemented in multi-agents
 distributed systems. At the best knowledge of the authors, argumentation-based trust
 models are not still investigated. The contribution is not only in the idea of using
 argumentation theory in trust, but also providing a list of trust scheme on which the

structure of an argumentation can be build, in the form of a distributed protocol. The trustee and the trustier may negotiate trust by using our trust schemes, rebutting arguments or asking for further evidence using our critical questions or support and defeat arguments with our defeasible semantic. In an argumentation dynamic, each party can be delegated and it is responsible for collecting the required evidence to sustain/attack arguments.

Other Model's features

We wanted to stress how our model is an example of model where the human notion of trust is central. We avoided any reductionist approach but, on the other hand, we wanted to produce a computable model. We content that this can be regarded as a contribution. Rich models, such as the cognitive ones or the original Marsh's models are often criticized for being only high-level pointers to computations that are too demanding or not feasible. Even if our model does not claim to make complex cognitive models computable, it is an effort in the direction of computing trust as a distinctive expertise, a complex phenomenon that calls for justifications and motivations. Our model conclusions is fully justified and the fact that the reasons for the computations are kept explicit and human-understandable; make it a transparent decision support tool.

1.7 Thesis Structure

The following three chapters' aim is to present the design of our computational model of trust based on presumptive and defeasible reasoning.

In chapter 2 we introduce the notion of trust from a social science prospective, describing a comprehensive definition of trust by Romano [Rom03]. In the second part of the chapter we introduce the computational notion of trust, where trust is defined as functions with specific properties and different representations methods. The chapter contains also our review of the state-of-the-art of the ways trust is represented and computed in current trust models. Our state-of-the-art review is driven by two goals: identify the ways trust is computed and underlie the defeasible nature of these trust mechanisms. We show how, despite of this nature, none of the mechanisms is treated as defeasible. The various computational mechanisms identified, adequately adapted to defeasible reasoning, inform the implementation of our model contained in chapter 4.

In chapter 3 we describe Presumptive and Defeasible Reasoning. We introduce the concept of non-monotonic logic, describing in details the two theories that are at the core of the definition of our trust model: Walton's Presumptive Reasoning [Wal96] and the semantic of Defeasible Reasoning as defined by Pollock [Pol01].

Chapter 4 describes our model design, built on the concepts and theories introduced in the previous two chapters. The chapter presents a theoretical design of the method, that starting from an analysis of the available evidence produces as output trust values, quantitative prediction of entities' trustworthiness. The model adapts Presumptive and Defeasible reasoning to a trust computation. At the end of the chapter we describe the main features and theoretical contribution of the model.

Chapter 5 describes the implementation of the model. The chapter contains computational tools used in the implementation and how they were used to compute our list of defeasible trust schemes. The chapter describes also our implementation of a defeasible reasoning semantic, essential to implement our key-idea of combining different trust evidence.

Chapter 6 contains our evaluation. We start by defining evaluation criteria directly derived from our research issue. Then, we describe the comparative evaluation strategy, which quantifies the efficiency of computed trust values by comparing them with a reliable and accepted trust metric produced independently. We then present a set of experiments conducted in the context of the Wikipedia project [Wik06c] and the online community *FinanzaOnLine.it* [Fin08]. In the conclusive part of the chapter we discuss results achieved. A section dedicated to open issues and future works closes the thesis.

Finally, chapter 7 contains our conclusions and a list of open and future issues.

Chapter 2 Trust from Social to Computer Science

Introduction

The aim of this chapter is double: to provide a notion of Trust as it emerges from its multidisciplinary study, sin order to investigate if our assumption of trust as a form of reasoning is compatible with social science, and introducing the formal notion of trust as implemented in Computer Science.

We largely adopt the recent definition proposed by Romano [Rom03], introduced with the aim of unifying previous works on trust focused on single aspects of this phenomenon.

The resulting definition is a notion of Trust as a *complex evaluation* performed by the trustier about the quality and significance of trustee's impact over trustier's perceived sense of control in the situation. The definition introduced is a compatible working definition that can inform our reasoning-based model.

After defining trust from a social science perspective, in section II we describe how practitioners in Computer Science formalized trust as a function. We focus on the way trust is represented in computational systems and the basic properties assigned to the trust function. In

section III we conclude our chapter by introducing the methodology and the aspects of computational trust.

Finally, section IV describes the actual landscape of computational trust models, essential for the definition of our trust schemes and section V focuses on the problem of evidence selection in trust, that, from the prospective of our reasoning-based model, we see as a part of the fundamental problem of how to initiate the arguments of such a reasoning.

2.1 Trust in the social sciences

2.1.1 Romano's ten defining characteristics of trust.

In this section we look at Trust from a variety of disciplines. The aim of this section is to collect conclusions of several authors regarding the nature of human trust, which should inform a computational model. This section is therefore not a comprehensive review of trust studies in social science, but rather we focus on major conclusions.

Literature about trust in social sciences suffers from lack of convergence. As trust is a fuzzy and somehow elusive concept, social science has analyzed this phenomenon with experiment in different discipline with narrow scope, ranging from psychology to business [Rah05]. However, it is still possible to underline basic definitions and properties.

Romano [Rom05] has recently performed an accurate study on the nature of trust, with the aim of clarify how trust might be defined and measured. Romano identified ten basic elements of trust, which are recurrent in many authors' definitions. However, no authors has a definition that encompasses all these aspects and, according to Romano, many definitions misunderstand the nature of trust, inserting in the definition factors that are antecedents or consequences of trusting, leading to an increasing confusions around this concept.

The ten partial definitions of trust, along with the common misunderstanding, are now briefly analyzed with references to other authors work.

1. Trust as an attitude. Trust is a personal attitude of the trustier towards somebody or something (the trustee). Saying that trust is an attitude it means that is a psychological event rather than behavioral. Trust is therefore not to be confounded or equated to actions like cooperation or risk-taking. Trust can actually occur without this behavior and vice-

versa, as confirmed by Gambetta [Gam00] and [Luh00]. Cooperation is therefore a potential outcome of trust. Trust is more appropriately an attitude, a subjective phenomenon that is defined by the psychological experience of the individual who grants it. The experience of trust is better characterized by the fought, feelings and behavioral intentions of an individual. From a computational point of view, the depth implication is that trust should be evaluated trough a kind of subjective reasoning rather than a deterministic model of a behavior.

- 2. Trust is social. Trust is an attitude towards a trustee in a specific context. Trust is social in the sense that is an interaction between trustee, trustier and context. Here the misunderstanding is about defining trust focusing on characteristic of only one of these three elements. Trust has been therefore related to trustier values or context's structure or trustee's attributes. These are more precisely reasons why an entity could decide to trust. Reasons to trust do not completely define the experience (and decision) of trust, but rather it is the perceived influence of such reasons, their assessment on one's outcomes and interests in a given situation.
- 3. Trust is versatile. Trust happens in many contexts and is granted to many different trustees. An issue is therefore if the nature of trust is a single construct that can be applied unchanged in different situations, or if there are many different typology of trust depending on situations. For instance, McAllister [McA95] identifies two type of trust: cognitive-based and affect-based, the first one present in task-based situation, the other in informal and interpersonal situation. Lewicki and Bunker [Lew95] identify three types of trust: calculus-based, knowledge-based, and identification-based.

In general, the idea is that trust *reflects* the situation. Romano considers situational factors as antecedents to trust, that influence the extent trust is experienced, but they do not represent different type of trust. Trust is considered a single construct, versatile in nature, which a trustier adjusts according to the situational factors. Trust is therefore independent from situational factors in the sense that is a generic construct adaptable to situations. A computational model should therefore include situational-dependent parameters, but it is allowed to keep the same kind of computation across multiple situations.

4. Trust is Functional. Trust is used by a trustier to obtain a sense of control in a situation in which it may not otherwise exist. Typically, the situation has uncertain outcomes, limited choices of action. Trust is a best estimate for making interactions decisions that promote desirable outcomes, and it is perceived as an instrument to have control over the situation.

The link between sense of control and trust should not be misunderstood. This relationship is usually assumed to be positive linear, such as trust represents a perception of heightened control, whereas lack of trust represents a reduction in the perceived control. From another point of view, many authors, notably Gambetta [Gam00], pointed out how meaningful trust has to happen in situation where the trustier is somehow vulnerable to the trustee's actions. In his classical definition trust happen in a context "in which trustier cannot monitor trustee's actions that can harm trustier". This point of view is opposite to the first one: the relationship with control and trust is positive in one case (you trust to have control), negative in the other (by trusting you lose control). Romano pointed out how the contradiction arises from a limited definition of trust in term of control. The functional nature of trust is so that, in both cases, the trustier has to obtain a functional sense of control that may be achieved through a willingness of unwillingness to be vulnerable. Therefore, trust is not how much or little control is perceived in a situation, but rather an assessment/evaluation that provides a sense of control to make productive interaction decisions.

- 5. Trust is hypothetical. Trust is future-oriented, and is about influence that has yet to occur. Trust, as pointed by Gambetta [Gam00], contains a prediction about another's likely behavior. Although a trustier's attitude may be based on perceptions of past influence, it is separated from them. Again, past history is only a possible motivation to trust. Since trust is hypothetical, one's expectations regarding future events are a crucial component of trust, and trust is therefore a future-oriented sense of control in accordance with one's expectations. Future rather than past is the ultimate reason to trust.
- 6. Trust is emotional-oriented. Trust is an emotional evaluation of influence. Not only the trustee predicts what influence might occur in a given situation, but he attaches personal feelings to the outcome of such influence. In this sense, trust is a sense of control related

to the personal feelings and values of a trustier.

7. Trust is goal-oriented. Trust is a perception of influence in relation to one's goals, Trust is again functional, and trustier are inclined to behave in a way that promotes desirable outcomes. Trust has therefore the function of giving a goal-oriented sense of control (see Castelfranchi and Falcone [Cas98]).

The last three partial components set up the basis to quantify trust. The idea that trust is quantifiable is common to many authors, like Marsh [Mar94], Castelfranchi and Falcone [Cas00]. As stated above, trust is versatile, in the sense that is a single construct that is used in different situation. Trust levels vary not because what is experienced is different, but only the degree to which it is experienced. The final three components are:

- 8. Trust is symmetric, in the sense that trust and distrust are not separate concepts but a different sign (+/-) that the trustier assignees to the influence of assessment, a judgment about whether influence is positive or negative. Note how symmetric is not used by Romano as a property of a mathematical relationship.
- 9. Trust is incremental, it is perceived at different levels with different magnitude. According to some authors, "low trust" refers to a negative, suspicious perception of influence, while others imply that "low trust" is a lack of trust. As a result, high distrust and low trust could be equated, leading to confusions and misunderstanding. In this thesis we clear this possible confusion by following Romano classification: trust is a positive assessment that can be perceived with different size, but still positive. Distrust is vice versa a negative assessment. This implies the existence of a neutral midpoint called *ambivalence*.
- 10. Finally, *trust is conditional*, i.e. it is perceived with different intensity. This dimension of trust gives the strength of trust, the extent to which a trustier is adamant in its perceptions of influence. As an example, the statement "I am very much convinced that this decision will slightly limit our outcomes" is a statement where trust is negative (*limit our outcome*), with low magnitude (*slightly*) and high strength (*very much convinced*).

From a computational point of view, characteristics 8,9,10 lead to treat trust as a quantifiable, orderable and fuzzy object.

Subjectivity vs. objectivity of Trust

In order to complete our review of trust properties from social science, we look at its *subjective* nature, discussed by all the social scientist that impacts computational models of trust. It is out of discussion that trust is something experienced subjectively, and its nature can be completely understood only by taking the personal perspective of the trustier. This subjectivity has strong implications in the way trust should be computed. First of all, the subjectivity nature of trust poses questions about its transferability: if trust is subjective, this means that it can't be transferred straightforward, without understanding the possible differences of view between two trustiers. Any attempt at objective measurement can dangerously mislead practitioners into thinking that the value is transferable and used by another trustier, which is not true for trust. In other words, trust is not transitive, which has also been formally shown in [Chr96]. As Luhmann puts it [Luh00]: *Trust is not transferable to other objects or to other people who trust.* To say that one trusts another without further qualification of that statement is meaningless, but it is meaningful when qualified.

Not a single computational mechanism of trust would survive by assuming that trust does not exhibit some objectivity or by assuming that trust is no more that a set of subjective policies. In order to exist, trust must be granted in a social context where trustier and trustees are interacting on the base of some *common (presumed) understanding*. Even if clearly subjective, it couldn't exist as a purely subjective concept, and it is its *weak objectivity* that makes it possible to happen, based on an agreement on intentions, goals, benefits, expectations among the parties. Trust is social and exists exactly when it is transferred among individuals (that therefore assume to have a common understanding) allowing social interactions to occur, as the glue of the society [Gam00]. Trust is subjective, but actually it wouldn't exist without its degree of objectivity, since it is a human society phenomenon rather than an individual one.

We think that the study of trust, even in a computational context, is the study of its degree of objectivity, how trust evolves and spreads supporting social interactions. The degree of objectivity may vary; it can be the results of a long emerging process or an artificial agreement. The subjectivity of trust, or its lack of objectivity, must be clearly reflected in a computational

model able to deal with ignorance, uncertainty and plausibility, but also a model that does not avoid to define the recurrent *contents* of trust as a distinct and objective expertise.

2.1.2 Definition of Trust

We are now ready to introduce a definition of trust. Our work is structured over Romano [Rom03] definition, where A is the trustier and B the trustee. Trust is a subjective assessment of B's influence in terms of the extent of A's perceptions about the quality and significance of B's impact over A's (potential) outcomes in a given situation, such that A's expectation and inclination toward such influence provide a sense of control over the potential outcomes of the situation.

The above definition appears the most comprehensive and generic at the moment, summarizing many aspects of previous definitions. For a comparison, we recall the classical Gambetta definition that had greatly influenced computational trust studies [Gam00]:

Trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently or his capacity ever to be able to monitor it) and in a context in which it affects his own action.

We note how Gambetta definition stresses the notion of trust as subjective probability, close to the notion of assessment of Romano, and the *hypothetical* nature of trust as a way to assess future actions. Gambetta stressed the notion of vulnerability of the trustier in order to have a definition of trust that in Romano is not a necessary condition. A missing dimension of trust is also the role of the environment.

Correct but partial view of trust is also present in other significant definitions. Deutsch's definition [Deu62] stresses the nature of trust as a cost-benefit game or a fulfillment of moral roles; trust as a collection of different constructs is present in MacAllister's work [McA95]. For a detailed discussion of these works we remind to [Rah05].

The aim of giving a definition of trust from social sciences was to set the basis of a computational model aware of such definition and compatible with it. We have interpreted some issues identified by Romano to be used to inform our model.

First, trust is an assessment of influence performed by the trustee. This evaluation is done

over some inputs, some reasons to trust or distrust, but it is the *perception* of these reasons for the trustier that really makes a trust decisions. The trustee should convince himself of the decision is going to take. Therefore it seems sensible to model Trust as a kind of *reasoning* where understanding the plausibility and the significance of the evidence given in a specific situation is the core dynamic of the process.

Trust is not to be confused with the reasons that could sustain a trust decision such as the ability of the trustee, past experience or reputation. Consequently, it is limited to derive trust in a deterministic way starting from the presence of such evidence. It is the reasoning performed by the trustee above the plausibility and significance of this evidence, mixed with its disposition and expectations that grounds a trust decision. A trust computation is strongly an assessment of plausibility rather than a prediction. This concept is confirmed by Luhmann [Luh00]: *Nor is trust a prediction, the correctness of which could be measured when the predicted event occurs and after some experience reduced to a probability value.* Social events cannot be analyzed effectively with probability only, since they are not well-defined repeatable experiments.

Any computational mechanisms of trust appears useful for giving reasons whether trust or distrust, but it is limited without a method to assess their plausibility and significance, and a way to combine and compare different finding into a more consistent and sound decision.

Second, Romano's analysis stresses the *contradictory* nature of a trust decision, based often on "different goals and conflicting perceptions" [Rom03]. Therefore a computational model of trust should include a clear strategy of conflicts resolutions.

Third, the three components of trust (8,9,10) put the basis for a clear quantification of trust judgments that will inform our model.

Finally, even if trust is a subjective assessment, this do not preclude that some *weak objectivity* among entities and environment is assumed. In this sense, the ten defining properties of Romano give us an accurate framework of *generic reasons* to trust that will inform our model. In the light of this evidence, we believe that our idea of modeling trust as a defeasible reasoning over plausible evidence is an adequate representation of the underlying theory.

2.2 Trust in computational models: definitions and properties

In this key-section we review how the notion of trust has been formalized as a computational concept. A trust model encompasses a formal notion of trust and a set of

computational mechanisms. The formal notion defines trust as a quantitative concept, with a correct representation, properties and dynamics.

The aim of this section is to introduce the different properties and representations of trust, discussing how they are adequate to represent key concepts as plausibility and uncertainty.

2.2.2 Formal Model of Trust: trust function, values and properties

2.2.2.1 The Trust Function

Starting from the seminal work of Stephen Marsh [Mar94], the vast majority of authors formalize trust as a relation (usually a function) F_t .

Marsh defines *basic trust* as the general trusting disposition of an agent, maturated in all its life. Following the notation of Carbone and Nielsen [Car05], we define P the set of all principals that can act as trustee or trustier, and T the function's image, representing the set of all trust values. The starting basic trust function is therefore:

$$F_t(P) \to T_v$$
 (2.1)

where P represents a principal or agent. The definition is not trivial and should not be underestimated: it actually states that trust can be quantified trough a function.

Formula 2.1 is the minimal definition of the function F_t, which takes as only argument the trustier and returns a trust value independent from a specific trustee.

The function domain can be extended by consider what Marsh called *General Trust, or Trust in Agents*. F_t is therefore defined on the Cartesian product of the trustee and the trustier. Therefore:

$$F_t(P,P) \to T_v$$
 (2.2)

Trust is a function that assigns a trust value t to each pair of principals a and b. Formula (2.2) adds the concept that trust in some degree depends on the specific individual trustee.

As a quantitative value, trust can be *ordered*. Again, Carbone and Nielsen [Car05] showed how trust forms a complete lattice of trust values. By introducing an order \leq among the image of the function F_t we imply that trust has different levels that are comparable with each others, so we are allowed to say that a is a more trustworthy entity than b.

As formalized by Carbone and Nielsen [Car05] and noticed by all the authors, a trust value that a principal a assigns to a principal b may require to consider others' opinions on b, for example from principal c and d. As an example, bank b may trust customer c only if bank b' has enough trust value in c. This is a common situation in open and collaborative environment. In order to

accommodate this situation, we extend the definition of the trust function:

$$F_t(P, P, T_v) \rightarrow T_v$$
 (2.3)

F_t can be further generalized as a function from a group of principal to another group of principal. This allows accommodating the notion of *reputation* and *voice* [Pin07], which can be defined as a particular trust value that a group of principal assigns to an individual or group. Therefore,

$$F_t(2^p, 2^p, T_v) \rightarrow T_v$$
 (2.4)

Where 2^P denotes the power set of P. Formula 2.4 asserts that trust is a function of a group of principal, giving to it a degree of objectivity or at least of consensus.

Trust is also situational, that means that an agent a trusts agent b in a specific situation s. If S is the set of all the possible situations, the function F_t can be modified as

$$F_t(P, P, T_v, S) \rightarrow T_v$$
 (2.5)

a definition common to Marsh and to the Secure trust engine [Cah03]. Marsh [Mar94] adds further details to this concept, qualifying a situation with the *Importance I* and *Utility U* perceived by the trustier. The proposed formula is:

$$T_{x}(y,\alpha) = U_{x}(\alpha) \times I_{x}(\alpha) \times \widehat{T_{x}(y)}$$
 (2.6)

Where α is the specific situation and I and U are the importance and utility of the situation α as perceived by the trustier x.

Note that in Marsh's formula the impact of the situation over the final trust value (the product UI) depends solely on entity x perception of utility and importance of α , not on y. The dependency of T_x from y is contained in the last term of the formula, that summarizes all the past trust value that entity x collected about y in the situation identical or similar to α . The meaning of formula 2.6 will be discussed further when looking at the game theoretical approach to trust.

Many authors consider trust an evaluation performed in the light of evidence. Josang wrote how "in the today global computing, trust must be based on evidence and knowledge" [Jos96]. Evidence-based Trust covers a large part of trust studies [Cah05], where the word evidence is any piece of knowledge or fact used in the computation of a trust value. In order to make explicit this dependency, we introduce in the trust function the argument *E*, representing the amount (set) of evidence on which trust is evaluated. Therefore:

$$F_t(P, P, S, T_v, E) \rightarrow T_v$$
 (2.7)

Note how trust values transmitted from other principals are evidence themselves, but for the

purpose of our trust systems classification we prefer to make the division explicit. Since trust is dependent on evidence, Carbone and al. [Car05] defines another ordering over principals defined on the amount of knowledge (evidence) a trustier holds about a trustee. Therefore, A may trust B and C at the same level but it may have more knowledge of B than C, fact that can be taken into account in a final trust-based decision.

Finally, trust is not a constant value, but a dynamic one that changes as the interactions go on. Thus, we should define trust as

$$F_t(P, P, S, T_w, E, t) \rightarrow T_w$$
 (2.8)

where *t* represents a timestamp.

Note that there are two types of time-dependency. The first type is due to the time dependency of F_t arguments, such as the set of evidence E or the transmitted trust values T_v . The trust function F_t modifies its value when new evidence or interactions are available, so that at time t_1 and t_2 the value may different because of new knowledge acquired.

Second, F_t may have a time dependency independent from new interactions, but intrinsic to its nature. This could be due to memory constraints (the trustier *forgets* past interactions and evidence after a while) or a direct dependency of F_t on time (for example, trust might decade after a long time since it is considered not valid anymore, even without new evidence). For example, the formalization done by Marsh [Mar05] of the concept of *forgiveness* is an example of intrinsic time dependency where a low trust value increases (is forgiven) without new interactions.

Issues related to trust's time-dependency include: the dimension of the trustier memory and the amount of time sufficient for modifying trust values.

The type of memory is also an issue. Memory can be defined as an *integral* operator that in one value encompasses all the past history (Marsh wrote how *a single trust value has the power of encompassing a great amount of knowledge* [Mar94]) or it can be a repository of all the evidence for more detailed further computations.

The *strudel* model by Quercia et al. [Que06], has a trust aging model. In this model trust is defined on n levels, and each of them has assigned a probability representing the likelihood of the entity's trustworthiness to be in that interval. The *aging* process is carried out by decreasing the trust probabilities that are greater than to 1/n and increasing those that are below 1/n, so that trust probabilities naturally tend to a uniform distribution over the n values. The method tends to forgive both good and bad past interactions after a proper amount of time.

In conclusion, computational trust defines trust as a function from a trustier to a trustee, or from a group of trustier to a group of trustees, that depends on time (intrinsically or trough new evidence), it can be ordered (or at least partially ordered); it is situational, evidence-based. We are allowed to write: agent A trusts agent B in the situation S a time t, in spite of the evidence E and a reported value of trustworthiness T from agent C.

How the trust value T_v is represented, i.e. the image of the function F_t , is the topic of the next section.

2.2.2.2 Trust Representation

As a function, trust has a value. The representation of this value varies from author to author, but few categories can be identified. These differences are more than mere choices of numerical representations, but they reflect the notion and property of trust underlying each method, and the range of applications it was intended for.

A basic representation represents trust as a one-dimension scalar. In the simplest form this can be a Boolean value, 0 or 1. Authors who decide for this representation consider trust something that is present or absent, it cannot be experienced at different levels. We consider a Boolean representation limited to specific trust-based applications, typically (hard) security-oriented situations, where the trustworthiness of an entity is associated with the presence (or absence) of a valid trusted certificate. In these applications trust is reduced to verify that an entity posses a particular object.

Other representations include a continuous interval of values, such as the real interval [0,1], or a discrete set of values, where trust is something quantifiable at different levels. A largely used definition based on a discrete set of value is a 5-level representation where labels are attached to values such as *very untrustworthy* linked to the value 0 or *very trustworthy* to the value 4. Other authors, like Golbeck [Gol00] propose a scale from 1 to 9. In the trust model of Abdul-Rahman and Hailes [Adb00] trust has a representation with 4 discrete levels corresponding to the labels vt=very trustworthy, t=trustworthy, u=untrustworthy, u=very untrustworthy.

It is interesting to notice if the interval chosen includes negative values. In Marsh, for example, a trust value varies from -1 to 1. This representation can accommodate both the notion of *trust*, *distrust* and *mistrust*. We recall how *distrust* is defined as the lack of trust based on some evidence or interactions, where *mistrust* is a diffidence not based on evidence. A value of 0 in this

scale usually means that the trustier does not have any opinion about the trustee, and therefore any positive or negative trust value associated with him. Note how the trustier may have collected a lot of evidence about the trustee, but still no clue about its trustworthiness. A value of -1 means that the trustier absolutely distrusts the trustee, while the value +1 means absolute (or blind, as many authors refer to) trust.

A conceptually different representation of trust is the one based on probabilistic distributions. A trust value is not represented by a single scalar, but as a probabilistic distribution over a set of values. The set of values could be one of those describe above, and for each element of the set a probability is associated. The probabilistic distribution partially resembles a fuzzy-set approach. It is able to represent the fuzziness associated with a trust value (the trust value may vary in a range rather than being a fixed value), and the compactness of the distribution gives an information about the certainty of the value. Two different trust values are compared using a fuzzy-matching that evaluates by evaluating the magnitude of the overlapping area between the two distributions.

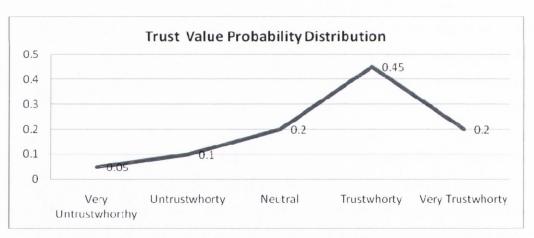


Figure 2.1 Trust value Probability Distribution

Representing the Uncertainty of a Trust Value

As trust is obviously an imperfect evaluation, it is affected by uncertainty. We have just seen how a probabilistic distribution representation contains information about the degree of certainty of a trust value, represented by metrics such as the standard deviation.

Many authors explicitly represent a trust value as coupled with a value of uncertainty. The image of the function F_t is therefore not only T_v but the couple T_v and the certainty value c. In the CertainTrust model, developed by S. Ries [Rie08], a trust value t is coupled with a certainty

value c to form an opinion o:

consider a value certain.

$$o = (c, t_v) \tag{2.9}$$

If the certainty value c is high, the trust value is a good estimation of the real trustworthiness of an entity, while if the certainty value c is low, the value of t may be affected by great variations. Quercia [Que05] provided a way to compute the confidence of a trust value, based on the variance of past interactions outcomes and their number. In the ReGreT model [Sab01], a context-dependant value has been introduced to define how many interactions are required to

Many models do not use a single numerical value for describing trust, but bi-dimensional or n-dimensional representations.

An extremely successful representation of trust is the subjective logic of Josang [Jos96, Jos05], a representation reflecting the human notion of belief and able to express uncertainty. In his model, every opinion is expressed as a subjective logic triple (b,d,u) where b represent the belief regarding the opinion, d the disbelief and u the uncertainty. Each element in the triple is a real number in [0,1] with the following property:

$$b + d + u = 1 \tag{2.10}$$

An opinion regarding a binary situation (such as do I trust you or not) is expressed as a probabilistic estimation (represented by b and d) with uncertainty u. When u=0, b+d=1 and the subjective logic without uncertainty becomes the probability calculus. When b=0 or d=0, it is reduced to binary logic.

Josang applies his subjective logic to trust scenarios, especially on-line rating system, where the triple (b,d,u) becomes (trust, distrust, uncertainty).

Formula 2.10 define the region of possible value of (b,d,u) as a triangle those vertexes are situations where two elements of the triples are null. The triangle of possible values is depicted in figure, 2.2. The base of the triangle is the region where uncertainty is null called region of the dogmatic beliefs. The point ω , for example, is in a region close to the belief (trust) vertex, representing a trust value with higher degree of belief and small uncertainty.

Josang defines new Boolean operators for trust propagation and aggregation as described in the next section.

In the Secure trust model, an evidence-based trust value is 2-dimensional, described by a couple of non negative integer numbers (n,m) representing the number of positive and negative evidence available so far. This representation maps to subjective logic by interpreting n as a value of belief

and m as a value of disbelief.

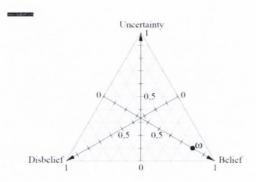


Figure 2.2 Subjective Logic Triangle

Uncertainty is given by looking at the difference between m and n (note how uncertainty is high when m and n has a similar value, low when the gap between them is higher. Therefore

$$b \propto n$$

 $d \propto m$
 $u \propto 1 - |m - n|$ (2.11)

In general, each representation carries a certain amount of information, and a possible conversion between two different representations implies a loss of information. Pinyol et al [Pin07] performed an analytical study of trust value conversions. The authors consider four common trust value representations: *Boolean, Bounded Real* in [0..1], *Discrete Sets* on 5 levels and *Probabilistic Distribution* representation. The authors proposed a conversion between these four representations and a factor to take in consideration the uncertainty involved in the conversion. For example, to convert from a Boolean representation to a bounded real representation, a Boolean value of false could be converted in any real number <= 0.5 and vice versa for a Boolean value of true. The target entity that is using a real number representation does not know how to place the converted value, and this represents the uncertainty involved in the process. In table 2.1 we report the quantification of the uncertainty associated with this conversion, computed as the entropy of trust as a random variable. High level of uncertainty are associated when a conversion is from a simple to a richer representation, while no uncertainty is associated with the opposite conversion.

Table 2.1 - Uncertainty in value conversion

| во | BR | DS | PD |
|----|-------|----------------------|--|
| 0 | 1.29 | 5.64 | 21.19 |
| 0 | 0 | 4.32 | 19.89 |
| 0 | 0 | 0 | 15.55 |
| 0 | 0 | 0 | 0 |
| | 0 0 0 | 0 1.29 0 0 0 0 | 0 1.29 5.64 0 0 4.32 0 0 0 |

In conclusion, in this section we reviewed how trust values are represented. Among the different representations, some authors propose a discrete variable set while others a continuous interval. Each representation has its reason d'être. The presence of a discrete level scale tends to standardize the values, especially with the use of labels, while a continuous representation may be more appropriate for computational applications and it carries more information. As Sabater writes in his trust models review [Sab05], a representation of trust with discrete levels suits more reputational on-line models, where input is collected interacting with humans, while a continuous approach is more suitable for probability-based trust computations.

In order to cope with the uncertainty inherent to a trust judgments, probabilistic distribution, fuzzy-set like or multi-dimensions trust value where have been proposed.

Models that represent trust with uncertainty or confidence levels can be seen as a first step in the direction of considering trust as a question of a subjective evaluation of a set of evidence, requiring a representation modeled after the human notion of belief.

2.2.2.3 Properties of the Trust function

Trust, as a function, shows some properties. Some of these are used as computational mechanisms to update and propagate trust values as described in the next section; some others better define the nature of trust itself.

Trust is **not symmetrical**: entity A can trust B but B may think that A is a very untrustworthy entity. Even if this statement is trivial, many social scientists underline how, in some way, there is a tendency to symmetry in trust, especially when an entity knows how another judges him. This means that A's trust in B influences how B trusts A. Even if this is not a necessary symmetric property, so that any time A trust (distrust) B then B trust (distrust) A, there is a relationship that cannot be neglected. Rotter and Deutsch [Deu62] show how people are more

likely to trust another person when they know that he trust them. As Marsh pointed out, if we know that a person trust us a lot and it has a high reputation of us, we can reasonable think that, in case we need his help in a situation involving trust, he will try not disappointing us, giving an evidence (of course not ultimate but plausible) that he can be trusted. Again, we see how trust is always a question of collecting some potential evidence that should be evaluated carefully.

On the contrary, if we know that an entity does not trust us, our inclination towards him could be suspicious and trust not easily to be trusted.

This concepts is partially modeled in [Mar94], that defines the value of trust of entity x towards y as a function of many factors including trust value of y towards x.

We therefore reject the naive vision of trust as a totally asymmetric in nature, as stated by Golbeck [Gol02], while we adopt a vision where other's trust in us may have considerable influences over our trust in them.

A connected property is the **reciprocity** of trust: if A helped B in the past, B will be more likely to give back the favor, supporting an evidence to trust B.

Trust is also **not reflexive**: it may be the case that an entity does not trust itself in respect to some situations.

A second class of properties refers to situations where trust values coming from different entities are in some way compared or **aggregated**. All these properties are highly discussed since they have to face the **subjective** nature of trust and its **transferability**.

A first concern is whether or not trust can be aggregated, i.e. if a trust value can be derived by a collection of trust values regarding the same trustee/situation but coming from different sources. This property is the basic assumption of any recommendation systems. Trust can be of course aggregated in the context of a single agent that has several past experiences about an entity, but what about opinions of different agents?

Authors agree about the possible aggregation of trust value, but many of them underline how the uncertainty coupled with the final trust value should be managed. They all agree about the need for a mechanism able to translate trust value in order to be aggregated, mechanism that requires an understanding of the trust models of each agent. Many authors aggregate trust without a mechanism to cope with uncertainty, leaving the final trust value meaningless.

Josang in his subjective logic proposes an operator for aggregating different trust values. Given two opinions about the same entity, expressed by two triple (b_1,d_1,u_1) and (b_2,d_2,u_2) the aggregation operator defines the resulting opinion (b,d,u) as:

$$(b, d, u) = (b_1, d_1, u_1) \oplus (b_2, d_2, u_2)$$
 (2.12a)

$$b = \frac{b_1 u_2 + b_2 u_4}{k} \tag{2.12b}$$

$$d = \frac{d_1 u_2 + d_2 u_1}{k} \tag{2.12c}$$

$$u = \frac{u_2 u_4}{k} \tag{2.12d}$$

$$k = u_1 + u_2 - u_1 u_2 \tag{2.12e}$$

The consensus operator produces a *consensus belief* that combines the two separate beliefs into one. Intuitively, the consensus opinion of two possibly conflicting opinions is an opinion that reflects both opinions in a fair and equal way.

Beliefs and disbeliefs of entity A are mitigated by uncertainty of entity B and vice versa (2.12b and 2.12c), while the global uncertainty always decreases to a value less than u_1 and u_2 . If the two opinions have both u=0, i.e. agents are perfectly sure of their beliefs, opinions cannot be aggregated (since k=0 in 2.12e). These are called dogmatic opinions by Josang.

Close to aggregation is the problem whether trust is **transitive**, and the same conclusions hold. Trust is not transitive, but, under some conditions and the study of its plausibility, remains a central computational mechanism for trust. Due to its importance as a computational mechanism, we will analyze it in a dedicated section in the chapter V.

Another issue is regarding the **situational** nature of trust. Authors agree that trust is situational, but on the other hand is true that trust is often "transferred" among situations. The key-issue is whether what we know about the trustee in situation s_1 is valid in situation s_2 . It is out of doubt that many trustee's characteristics, like honesty or commitment, that are evidence for a trust evaluation, are not situational, while things such as competency are, since situational-specific knowledge could be more important than general-purposes skills. For example, a brilliant CEO of a car industry could likely be a brilliant CEO of an airplane company, while a brilliant surgeon cannot be trusted in repairing cars. Saying that trust is situational cannot be reduced to the fact that, if s_1 is different form s_2 , therefore trust cannot be transferred among situations, since an analysis of s_1 and s_2 similarities could make it a plausible computational mechanism.

A study about how to transfer trust among situations and under which conditions it can be considered plausible has not been done yet, even if proposed by the *T3* research group [T3G07]. Mechanisms have been proposed to consider a degree of similarity between two situations and

used it to modulate trust values. Moreover, almost any trust models have the notion of basic trust that recognizes a situation-independent nature of trust.

Trust is **self-reinforcing**, as identified by many social scientists. Rotter and Deuscht [Deu62] stated that "trustworthy people are more likely to trust, while untrustworthy people generally trust less". Marsh wrote how "if you are capable of trust, you will trust, if you are not capable, you will consider the other not capable too." The author commented how this means that the trust function has a dependency from our general disposition of trust, forming a kind of accelerated cascade effect where high value of trust will make the entity likely to trust more and vice versa, bringing the entity towards the opposite position of an optimistic or a pessimistic vision of the world.

We wonder now if F_t is a **continuous** function and if it shows some degree of **regularity**. Many trust models, due to the way they aggregate evidence, result in a trust function that shows a **slow-to-react** and regular behavior. As an example, let's consider the simple eBay-like feedback systems. Since the new total trust value has a dependency on past values, F_t shows a behavior similar to an *integral* function, therefore continuous and regular.

Since a new trust values have a (partial) dependency on past ones, many authors considers F_t behaving as a continuous function that doesn't show discontinuity of sudden change of values but where changes are limited.

As a central issue of this thesis, we generally disagree about the smoothness and the continuity of the trust function, but rather F_t could present sudden discontinuity, even of great magnitude, that reflects the fact that trust may dramatically changes forgetting, collapsing or erasing all the previous history. A computational mechanism able to deal with this situation will require a new aggregation strategy we identified in the argumentation theory paradigm.

Finally, in some trust models like Quercia's Strudel [Que06] F_t exhibits a kind of **elastic memory** that reduces high trust values and increases low trust values, so that after a certain amount of time – in absence of new evidence – the function tend to an initial fixed value.

In conclusions, in this section we identified a generic pattern in the discussion of trust properties: it is accepted that trust does not show many mathematical properties, but none of them can be completely discarded, since considering trust having those properties is actually more useful from a computational point of view and closer to the human trust than the opposite.

Trust seems to show these properties in a weak way, but their presence is actually essential for trust to exist. Trust is subjective, but actually it wouldn't exist without its degree of objectivity.

Trust is not transitive, but transitivity is used every day in trust. Trust is not symmetric, but who is not influenced by others opinions about him?

When it comes to translate these properties into computational models, much of the work is therefore the understanding of the defeasibility (plausibility) of specific trust properties, its weak objective side in the context rather than their application straightforward in trust computation, as it usually happens.

2.3 Computational Trust Solution: components and methodology

In this section we introduce the generic elements common to any computational trust solution and the terminology used for the rest of the thesis.

We start by analyzing a typical computational trust solution, referring to fig. 2.3, modeled after the high-level architecture of the Secure trust engine [Cah05].

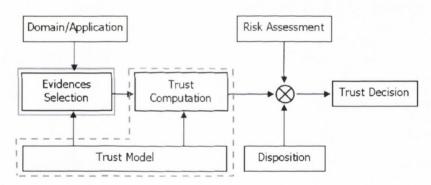


Figure 2.3 A computational Trust Solution

A trust-based decision in a specific domain is a multi-stage process. The first step is the identification and selection of the appropriate input data. These data are in general domain-specific and identified trought an analysis conducted over the application. We refer to this process as *evidence selection* and the inputs used to compute trust as *trust evidence*.

Evidence selection is driven by an underlying *trust model* that contains the notion of trust on which the entire system is centered. A *trust model* represents the intelligence used to justify which elements are selected as trust evidence, why some elements are selected and other discarded, and it informs the computation over the selected evidence. A *trust model* contains the *definition* of the notion of trust, its *dynamics* – how it evloves over time and with new evidences -

and the mechanisms of trust used in the computation.

A trust model can contain a rich human notion of trust, as in the cognitive models where trust is defined as a mental process whose ingredients are identified and investigated.

Other approach assigns a marginal role to trust by encoding limited *trust intelligence*. Usually these models are easier to be implemented and computed. Their main contribution is not focused on investigating the ingredients of trust, but limited to the formalization of a contaner for trust evidence and a way to propagate this evidence – mainly by transitivity – while the *actual provisioning of trust* values and its justifications is delegated to some external infrastructure (users'feedback, digital certificates).

After evidence selection, a *trust computation* is performed over evidence to produce trust values, the estimation of the trustworthiness of entities in that particular domain. A trust computation requires the formalization of a computable version of those mechanisms defined in the trust model. Example of such mechanisms are the past-outcomes one, reputation and recommendation, but also temporal and social factors, similarity and categorization and so forth as descibed in Chapter 4.

A *trust computation* should be able to quantify the impact of evidence found, chose a *representation* for the trust values and define a strategy to *aggregate* different evidence into a global trust value.

For instance, a classical trust system use two set of evidence: recommendations and past experience. Each of them is quantified separately and then aggregated into a final value, usually with a linear combination.

A trust system seldom has all the above stages at the same level of complexity. Usually, a system with a limited trust model results in a well-defined and bounded overall system and viceversa. For instance, the trust model underlying a PKXI [Bal03] is limited to the fact that trust can be certified and propagate by transitivity. Evidence selection is supported by a dedicated infrastructure that defines *apriori* the structue and nature of the evidence – digital certificates and signatures. The computation performed is a retrievial process along a transitivity chain, a challenging process but again well-defined.

In a reputation systems the underlying model is the fact that trust can be obtained by sharing information. Evidence selection is delegated to users, while the formal representation of trust and aggregation of different recommendations can be a non-trivial –but well-defined – task.

When the underlying model is a cognitive rich notion of trust, evidence selection is a complex

task. The model offers usually high-level pointer to class of evidence that has to be mapped onto domain elements. Moreover, formalizing a trust computation of cognitive trust ingredients is a complex, soft and seldom unfeasible tasks. After the computation, the aggregation strategy reflects the complexity of the quantified evidence, that can be of different nature and encompassing concepts that calls for a more reasoning-based approach.

Finally, the actual trust decision is taken considering computed values and exogenous factors, like *disposition* or *risk assessment*. The term disposition is used here to identify the attitude of the trustier regarding trust independently from the specific trustee. For instance, a trustier could be optimistic, pessimistic or neutral regarding trust.

2.4 Computational models: how to compute trust

In this section we review the current landscape of computational models of trust. Our main goal is to describe how trust is computed and the underlying reasons of each model. The findings of our rewiev will be used to in the next chapters to define our list of defeasible trust schemes.

By looking at the current landscape of the trust models, there is a trend that few authors underline. Much of the work is dedicated to the definition of trust as shown in the previous section: its formal representation, its properties and a correct representation of values. Those elements are almost sufficient for *propagating* trust. Many trust reviews are unbalanced, giving a strong emphasis to trust representation and propagation. Here we focus on the way trust can be actually computed.

At first sight, it seems that trust computation is a matter of a probability-based prediction done over direct past experience of the trustee, or based on indirect experience, that time to time is labeled as Recommendations, Reputations, Ratings, Social Networks. The relying on transitivity completes the computational scenario. Risk assessment, sometime considered a complementary but disjoint factor, sometime considered an essential component of trust could be present in the computation.

Merging many authors' taxonomy [Sab05] we classify computational trust mechanisms in the following classes, each of them described by a dedicated sub-section:

- Direct Past Experience: computing trust on interactions history
- Indirect Experience: computing trust using others' trust computations

- Probability-based Trust: treating and predicting trust as a probability
- Game Theoretical Approach: trust as an utility game played by rational agents
- Collaborative filtering, categorization and similarity: transferring trust trough similarity and categories
- Risk: deciding to trust based on risk analysis
- Cognitive Models: trust as a mental model
- Ad-Hoc Heuristics for computing trust and monitoring-based approaches

2.4.1 Past Outcome using Direct Experience

This computational mechanism uses evidence that the trustier gathered directly from previous interactions to predict trustee's future behaviours.

A clear definition of this computation, and the correlated notion of trust, is the one produced by the research group Trustcomp.org: "trust is a prediction of future behaviour based on past evidences" [Tru04]. Due to the predictive nature of this mechanism, it could be catalogued under the class of probability-based trust computation discussed in the next section. This is the choice of many trust review such as [Sab05] [Rie08]. Anyway, we prefer to describe this mechanism alone, since a generic probability-based trust computation could not use direct experience and because direct experience can be implemented without using probability distributions.

There are many different incarnation of the past outcome paradigm, but they all share a common basic scheme. The central notion is that a trust value is computed using the outcomes of all the pertinent past interactions. The value is updated when a new interaction occurs, proportionally to the outcome of this interaction. Advanced implementation considers a memory factor and an ageing process for the past-outcomes collected. A detailed analysis is described in Appendix A.

The direct experience evidence is without any doubt the most used computational trust mechanism. Anyway, there is too much confidence in its applicability. In Sabater's [Sab05] review of trust model, direct experience is regarded as *the most reliable source of trust*. We agree with this statement, since by relying on direct experience an agent should be sure of its level of satisfaction, reducing the noise due to the subjective differences that may arise by using others trust value. But the direct experience is, as any mechanism, a presumption that should be tested and that could be not so easy to implement.

First, the feedback function F, necessary to understand the outcomes of an interactions could be not feasible and the outcomes not measurable and quantifiable. This implies that F is implemented in a too approximate way that can lead to incomplete or meaningless trust values. We note how usually many of these problems are by-passed relying on human explicit feedback. Other factors that should be considered before relying on this computational paradigm are:

- the number of outcomes, their frequency and their temporal distribution that, may invalidates the trust value that becomes also subject to high variance,
- the stability of the situation/agents that may change, invalidating past evidence
- what it is possible to known about interactions' outcomes, which could be not observable or partially observable, making impossible or highly uncertain their evaluation and the consequent trust value.
- External constraints out of the control of the trustee that affects the outcome of an interaction, that should not affect trustee's reputation

2.4.2 The Theory of Probability as a Computational Tool in Trust

Probability theory is used to compute or propagate trust value. This class of trust computations is more a meta-class: they implicitly rely on other mechanisms, typically the direct experience seen above, to collect the required inputs. Probability is therefore a *tool* to compute trust value rather than a model *per se*. The common assumption of these systems – and the reason for their success - is that trust can be represented and manipulated (i.e. predicted) as a probability distribution function that typically models the expected behavior of a trustee. Advantages are a clear (but limited) semantic meaning and effective computational tools, such as Beyesian Inference where probability becomes not only a meaningful trust representation, but it goes beyond offering also mechanisms for *updating* and *learning* trust.

It is often coupled with past experience of indirect experience, but not necessarily.

We may wonder if probability theorems are a valid method for manipulating trust.

The assumption underneath probability-based trust is that trust is somehow a phenomenon that can be predictable, once having collected enough evidence to define a probability distribution function. As we have seen, social scientists consider simplistic a vision of trust as pure probability-based prediction.

Critics of the method underline how the assumption of representing and updating trust as a

probability is a meaningful but reductive approach that hides mental components of trust. As Castelfranchi and Falcone wrote [Cas00], it cannot be underestimated the importance of a cognitive view of trust (its articulate, analytic and founded view), in contrast with a mere probabilistic and quantitative view of trust supported by Economics and Game Theory.

The author of this thesis partially agrees with this statement. When probability is only the quantitative representation and the mechanism chosen to represent a complex, even cognitive, model, the use of probability is justified and it can offer the richest set of tools available. The problem is therefore which information is present underneath the probability computation.

When the value of probability represents almost all the information available, and no reasons or arguments are attached to this value, the approach is limited and inappropriate to deal with trust. The tendency is nowadays to reduce the complexity of trust models in order to keep trust computations simple and values easy to propagate, but this approach is appropriate only in specific context and domain where it is possible to verify that the *reasons* behind trust can be omitted without losing much of accuracy.

Unsupervised Learning tool for predicting trust

In the extreme case the idea of trust as a prediction can be obtained by the application of unsupervised learning tools that associate entity's features with its trustworthiness, or with good or bad outcomes. Strong points of this computation are the domain-independent nature of the tools used, and the precise quantification and semantic of results.

On the contrary, the role of an explicit notion of trust is limited or absent. Trust is *implicitly* defined before training the system, by collecting and labeling an adequate set of expected behavior examples. We could say that all the trust work is outside the computation that is merely another computational tool able to recognize what is similar to trust. In cases where results are fuzzy and uncertain little can be deducted. These computations rely ultimately on similarity of situation where trust can be transferred. As Romano clearly pointed out, trust is not a prediction in the sense that several social factors make pure trust predictability a weak argument. Due to the plausible nature of this assumption, an implicit unsupervised computation can be not appropriate to deal with special cases and situations that call for a more explicit approach, in which trust dynamics and arguments are kept explicit in the computation. Implicit unsupervised techniques may be the correct computational tool in domain where the notion of good and bad outcome is well defined and the range of cases show a consistent uniform and homogeneous nature.

Moreover, the choice and justification of training cases remain the critical issue from a computational trust point of view. When this preliminary selection is not justified, the final computation will lack meaning. Many learning techniques or statistical tools like principal component analysis may become extremely useful tools for trust practitioners to get insight of systems dynamics, but their utility depend on the ability of making the right interpretation or the insight offered, that requires an explicit notion of trust.

2.4.3 Indirect Experience

The computational mechanism behind *indirect experience* has some analogies to the *direct experience* one. The trustier agent collects, from other agents, positive or negative recommendations about the trustee, and it processes them in order to take better trust-based decisions. The core idea is that opinions of other agents are useful information to predict trustee's behaviour.

In this mechanism trust information is condensed in a quantitative representation containing past experience, direct experience and contextual information. The focus of this mechanism is to correctly propagate trust. Therefore it is more a tool for propagating trust rather than computing. Propagating trust requests to define an operator that aggregates different opinions, called *discounting* operator or *aggregator*, and operator that aggregates values along a *transitive* chain.

Many issues make the computation of the aggregated value a more fallacious process than the *direct experience* case. If direct experience is considered the most reliable source of trust-related information, indirect experience is usually the most abundant and hard to process. When agent A collects an opinion about a trustee agent B, issues that make second-hand information complicated to process are:

- 1. the information is coming from another agent, therefore the difference between agent A and B should be known in order to estimate the relevancy of B information for A. If B has a strongly different concept of trust than A, the same action could be judged in many different ways, making the recommended value meaningless. If direct experience has a degree of objectivity (at least for the trustier agent), indirect experience usually does not.
- 2. the information is coming from another agent, which may have collected this from its direct past experience or from others' recommendations. Therefore, the information could

come from a source that is *n-hops* far from agent *A*, with a consequent *noise* affecting its value.

- 3. the information is coming from more than one agent, and information collected can be contradicting.
- 4. nobody guarantees that agent *B* is trustworthy in giving recommendations. Agent *B* can act maliciously and transmit a unreliable value

Problem 1 is the issue of recommendations semantic meaning, and considers every aspect that can make the transmitted value not meaningful for the receiving agent, even if both the parties are acting honestly. This includes different expectations, evaluation metrics, different amount of past experience, different trust models and so forth.

In addition, the problems 2-4 described above add to the computation the following parameters:

- 1. *the proximity of the information*: information from a recommender directly known should be treated different from information n-hops far
- 2. the quantification of recommender trustworthiness should be part of the computation. Note how this value is the trustworthiness of the agent as a recommender (that is generally different from its trust value).
- 3. a strategy *to spot outliers*' values and possible *malicious* recommender. If with direct experience there is no room for malicious data, with recommendations this problem is crucial.
- 4. a strategy to *aggregate* values received, that may contain contradictions and inconsistencies. Moreover, we need to discuss the following:
- 5. how direct and indirect experience are merged in order to produce a single trust judgments

Note how problem 3 and 4 are complementary and orthogonal; the first one deals with an "in depth" direction of trust (vertical), the other with a horizontal dimension.

Social Network

Social network wants to exploit information about nodes' social relationships in order to propagate trust value. The novelty of the approach, as pointed out by Sierra in [Ope03], is in the use of sociological information as an evidence for trust. The social relationship established among agents in a multi-agent system are a simplified reflection of the more complex

relationships established between human counterparts. Computational model based on social networks are largely dependent on transitivity.

Social network analysis relies on the *Small World* property that many network, notably the WWW, exhibits. In this are we mention the work done by Golbeck [Gol02], the FOAF project [FOA07] and the REGRET model [Sab01], all described in details in Appendix A.

2.4.4 Collaborative filtering, Similarity and Categorization

The common idea behind these mechanisms is that trust can be transferred among similar entity/situation.

Collaborative filtering assumes that ratings from agents with similar tastes are more accurate. Collaborative filtering classifies object using the ratings agents assigned to objects, and it differs from content-based filtering which classifies objects based on their content. Collaborative Filtering knows nothing about the items' true content and they rely on preference values, such as ratings, to generate the recommendations.

Relying on ratings allow a classification of a broader range of object that using a content-based filtering, that is based on automatic algorithm, are hard to interpret correctly, such as music, or videos. This explains the large use of such systems in online reputation systems.

The basic mechanism of a collaborative filtering is as follows. When a new user enters the system, it provides a number of recommendations, or a personal profile, that is stored in the central database. When a user asks for a recommendation, the system will match user's profile with those stored in the database, and it will suggest recommendations from similar users. The efficiency of these systems depends on the number of recommendations it receives, the number of users in the system and the level of details of each user profile.

Usually collaborative filtering are centralized solutions, but Wang [Wan03] proposed to extend the basic idea of collaborative filtering in a dynamic and distributed fashion.

Similarity & Categorization Tversky largely studied [Tve74] how similarity and categorization are cognitive mechanisms used to make judgments under uncertainty. In the context of computational trust, similarity has been investigated by several authors, like Ziegler and Golbeck [Gol00], while Castelfranchi and Falcone [Cas02] consider categorization a first form of trust-based reasoning. The plausibility of the trust factor increases if the category set is numerous and compact enough, i.e. with a small variance.

Finally, Sabater in [Sab05] describes how *prejudice* has been used as an evidence to compute trust. Prejudice, writes the author, *is the mechanism of assigning properties to an individual base on signs that identify the individual as a member of a given group*. Therefore, we consider prejudice a form of categorization since an entity is pre-judged on the basis of some features that associate it to a specific category or social group whose trust value is known. We note how prejudice does not have the negative connotations that has in many human societies.

2.4.5 Game Theory: the utility game.

In the Game Theoretical approach, as described by Sierra and Sabater in [Sab05], trust and reputation is the result of a pragmatic game with utility functions. This approach starts from the hypothesis that agents are rational entities that chose according to the utility attached to each actions considering others' possible moves. Action could be predicted by recognizing an equilibrium to which all the agents are supposed to tend in order to maximize their collective utility.

In the definition of trust by Romano described in a previous section, one component of trust is the *motivational* one that makes agents acting according to their interests. Even if a trustee agent has proved to be very trustworthy in the past, and there is evidence about its ability to accomplish a trust purpose, if he has lack of utility and motivation he cannot be trusted.

The Game Theoretical approach in trust can also be encoded in the design of the application. In this case, the application is designed so that trust is encoded in the equilibrium of the repeated game the agents are playing. Thus, for rational players trustworthy behavior is enforced. A more comprehensive description of game theory in trust is described in Appendix A.

2.4.6 Risk

Risk is a factor that cannot be neglected in trust computations. Different interpretations of the role of risk are possible, ranging from considering a trust-based computation a particular kind of risk-assessment to considering risk an exogenous disjoint factor to trust. Coleman [Gab02] wrote how the incorporation of risk onto a decision can be treated under a general heading that can be described by the single word trust. We definitely consider risk an argument in a trust-based decision, even if we keep it separated and we treat it as a datum. In line with many authors, we consider risk a more defensive concept that suits mainly security-based scenarios, focused on

assessing the likelihood of events and their attached benefits/loss, while trust is a more cooperation-oriented concept.

If trustee A and B have a comparable risk, a trustier agent will select the one he trust more and vice versa. Decisions can be made only on one of the two factors, but relying on both make the process more plausible and complete.

In our model, risk affects a trust-based opinion rather than be part of the actual trust-based computation. Risk therefore can revert a trust-decision, but it does not affect the trustworthiness of agents. It is an exogenous factor that logically is separated by trust, but essential in a decision.

2.4.7 Cognitive models and Trust Ingredients

A cognitive model of trust defines the mental processes, mechanisms and dynamics of trust. These models stress the nature of trust as a complex structure of beliefs and goals, implying that the trustier must have a "theory of the mind" of the trustee [Cas00]. Trust becomes a function of the degree of these beliefs.

Cognitive models consider trust something defined, with its ingredients and rules. The mental states that lead to trust another agent, as well as the mental consequences of the decision and the act of relying on that agent, are an essential part of the model. Cognitive models present a rich notion of trust, and reject the reduction of trust to a probability-based computation, that is seen as a simple and limited approach, as described by Castelfranchi and Falcone in [Cas99], whose model is discussed in Appendix A.

The ingredients of Trust

The existence of cognitive models allows us to speak about *trust ingredients*. According to these models, trust is made of recognizable components. In our trust reasoning graph (chapter III), trust ingredients play an important role: they are directly connected to trust or distrust, and they therefore represent the high-level arguments why an entity should be trusted. We wonder which these trust ingredients are. The work of Castelfranchi and Falcone just analyzed and the work of Sztompka [Szt00] are used to define our trust ingredients.

According to Sztompka, trust is composed of 7 factors: regularity, efficiency, reliability, representativeness, fairness, accountability and benevolence.

Cognitive model by Castelfranchi and Falcone suggest inserting competence and fulfillment.

Therefore the following are our trust ingredients:

- Fulfillment that measures the commitment of the trustee to the task assigned
- Competence, reliability and efficiency that measures the trustee's ability to deliver in a specific context
- Regularity that suggests the importance of the time distribution of trustee's activity
- Accountability that measures how the trustee is accessible and transparent in his actions
- *Representativeness* that measures the commonality between the trustee and the category he belongs too, or the trustee.
- *Fairness, benevolence, motivation* that measures some of trustee's features that can increase the sense of control over the situation by the trustier.

Our trust scheme, described in chapter VI, directly support one of this ingredients.

2.4.8 Computing trust over domain elements: Monitoring and Heuristics

Trust-based Heuristics

Trust can be the result of a computation over elements of the application under analysis by mean of some kind of heuristics or intuitions. We specifically focus on computations based on domain elements, since any kind of other evidence, from human feedback to digital certificates, can potentially be manipulated by a heuristic.

This class of fragmented computations represents an attempt to extend trust computation in the way we would do, and therefore the following discussion will be highly relevant.

Generally, the procedure requires identifying some elements of the applications that are considered to have a meaning for trust according to the author of the heuristic. The values of these elements are processed with an ad-hoc formula producing a trust metric. While the computation presented so far were based on some systematic methods where inputs and computation were defined, an heuristic derives more from an interpretation that is given to some domain elements in the context of a trust computation. Elements used are usually part of the application without any explicit trust-related scope.

Several heuristics have been defined over domain elements. This shows that there is *confidence* that these kinds of computation are effective. They show also interesting properties: they are *application-contained*, not requiring a dedicated infrastructure for trust computations. In spite of the numerous examples, heuristics are considered a second-class computation, based on

intuitions or insight that are highly subjective and target of critics. The point is their lack of systematicity and objectivity or lack of (explicit) justifications to questions like: Which heuristics should I use? Which elements should I use? Why? Which is the meaning for trust?

Example of trust-based heuristics can be found in Appendix A. Our model intends to create a more systematic ways to approach a trust computation over domain elements.

Monitoring the system

Trust can be computed by monitoring a system and interpreting its dynamics. Usually, some expected behaviors or characteristics of agents in the system are used as an indicator of their trustworthiness. Crucial to the correct monitoring of the system is the ability of interpreting what is going, that requires an explicit model of behavior.

The same discussion we performed for heuristics is still valid: the model used to monitor the system should have a clear trust meaning, able to justify why some elements are indicators of trust. Heuristics and monitoring-based are similar approaches where the central issue is the interpretation and the meaning assigned to some application elements. Monitoring is usually a more systematic method, but the problem of a strong domain-dependency of some models and the little number of existing models leaves room for the definition of generic models able to monitor larger class of system. As a good example of monitoring approaches, Appendix A analyses the trust model proposed by Carter.

The main idea behind the reputation model presented by Carter et al. [Car02] is that 'the reputation of an agent is based on the degree of fulfillment of roles ascribed to it by the society'. If the society judges that agents have met their roles, they are rewarded with a positive reputation; otherwise they are punished with a negative reputation. The roles identified by Carter are: Social information provider, content provider, longevity role, administrative feedback role, interactivity role.

Carter method contains similarity to our proposed methodology. Trust is deduced directly by monitoring the environment and identifying elements that could be useful in the computation. Elements are selected by mean of an explicit notion of trust, represented in Carter by the five expected roles that an agent should fulfill. The five roles are similar to the concept of our trust scheme, since they represent the generic arguments to trust. The five roles suggest which elements of the domain should be used: any element revealing that the agent is fulfilling or not these roles is trust evidence. The model is therefore justified process.

2.5 The problem of evidence selection

The starting point of any trust-based decision is the collection of available evidence. Without a strategy for selecting evidence, no trust reasoning can start. In this section we analyze how current solutions treat this problem: the kind of evidence used, the way these are collected, the justification of the process. Since this is essential for the definition of our method, evidence selection has been kept divided from the computational mechanisms described in section 4.1.

In literature, the term evidence-based trust has a broad meaning. Evidence is the input of the computation. Many authors [Sei06] [Cah00] encompass reputation, recommendation, certificates as evidence. We focus mainly on a subset of evidence-based trust computation, where evidence used has a justification *inside* the trust model and not delegated externally, since they are the more unexplored. We begin our discussion by analyzing the possible source of evidence used in trust computation in section 4.2.1, while in section 4.2.2 we review the method used to collect evidence.

2.5.1 Type and location of evidence

The set of evidence used in present trust solutions is diverse. We consider evidence any input of a trust computation, regardless the reason behind their usage, implicitly or explicitly stated. We divide the current set of evidence in three groups according to the nature of evidence that can be (i) part of a dedicated infrastructure, (ii) feedback of various nature (human-based, autonomic..), or (iii) identified over elements of the application where trust has to be computed. The main difference is the complexity of the problem that remains well-bounded and partially self-justified in the first two cases, unbounded and affected by subjectivity in the last case.

2.5.1.1 Ad-hoc evidence in a dedicated trust infrastructure

Trust computation may be based on a dedicated infrastructure explicitly added to applications. A classical example is a trust computation based on keys or digital certificate (see for instance [Bal03]). The PKIX is an independent layer that can be added to any distributed application where it provides not only security encryption but also the support of trust-based decisions, for example using digital certificate. In a typical scenario, a user has to trust a service or some information coming from a likely unknown entity that shows a digital certificate as

evidence of its trustworthiness. In this paragraph we discuss about the concept of trust in this scenario, we are interested in the nature of the evidence used.

In dedicated infrastructure evidence are *well-defined* objects with a defined life cycle. They have been designed as trust evidence and there is no need to justify their meaning for trust. In applications where a dedicated infrastructure exists, an entity is trusted if it can demonstrate to have a valid evidence. The hypothesis behind this set of applications is that valid evidence is equal to trust. This approach does not show for us any differences with a password-based approach and it is a distributed version of it. Since evidence are objects known *a priori*, the problem of evidence selection is therefore not present or reduced to the verification of defined evidences, which may imply operation of retrieving and searching over the infrastructure. The validity of evidence is usually certified by particular entities that are trusted unconditionally. Let's consider again a typical PKIX infrastructure with Trusted Third Party. Entity are trusted if they can show a valid digital certificate, that is signed by a TTP, that is trusted unconditionally by all the entity. TTP forms a hierarchical tree, where upper levels TTP act as guarantor for their child TTP. It is usually not known to the entity on which basis another entity obtained a certificate. Usually certificate has a temporal validity.

For application where is not possible to defined such trusted entity, entity may certifies themselves, as the case of PGP. PGP does not rely on the presence of a TTP and users trust other users if they can show a certificate that is validated by trusted users.

Again, what we want to underline is that trust is based on the possession of a valid object that has been created for that purpose, and whose trust-related meaning and life cycle are encoded in the infrastructure mechanisms and familiar to any entities using the infrastructure.

The propagation of trust value is usually done by transitivity. We note how the transitivity mechanisms, that implies the fact that trust can be transferred forming a chain of entities trusting each other, is facilitated from the presence of a well-defined concept of evidence and consequent trust, that are interoperated in the same way among different context. Evidence and consequent computation cannot be completely separated.

2.5.1.2 Using feedback as evidence: humans' ratings, outcomes, trust values,

The evidence used is represented by the explicit evaluation of an interaction or agent. This situation has strong similarity with the previous one, especially when the process is entirely

delegated to users: evidences are well-defined object that are part of a dedicated infrastructure, typically a recommendation system.

Humans rating and Outcome evaluation

In this situation evidence is represented by feedbacks and ratings that humans insert in a system that process and aggregated them. In these systems we rely on users' collaboration, we delegate the gathering of evidences (user's ranks) to them.

The trust meaning of a rating is explicit, even if not always semantically appropriate. Examples of successful trust solutions are feedback and rating systems, Social Networks, the FOAF infrastructure and their extensions [Gol00] where users (or agents) explicitly define what they trust or not.

Using others' trust values

Trust values can be used as evidence for other trust value. When trust values are coming from users, the situation is analogous to the previous one.

Using trust values as evidence for computing other trust values is a solution that ultimately delegate the evidence selection process to a previous computation external to the process. The use of trust value is common to various classes of computational mechanisms: indirect experience, solutions that propagate trust (social networks, transitivity) and solutions that need to be explicitly trained. In the first two cases others' trust value are the input of a computation, while in the last case the system to be trained with a set of known trust values. The problem of how the trust values are generated starting from evidence remains external to the system, exactly like a human feedback.

2.5.1.3 Application elements as trust evidence: with or without trust

In this situation, application's elements revealing the presence of Trust are used as trust evidence. Here evidence is not defined a priori and its selection becomes a complex problem. No user intervention is required; no dedicated infrastructure to collect evidence is added to the application and the computation has to follow different ways than the feedback-based loop described in the previous section.

The idea is that some application elements carry trust-related information, as result of

interactions going on in the systems that have attributed to them an implicit trust meaning. Some elements may resemble or act like an implicit version of a trust mechanism and some others may have been recognized by users as having a role for supporting decisions where trust is needed.

This trust-related meaning could be perceived by the users with different degree, and in general is a presumption whose plausibility can be argued.

In literature, many example of trust computations performed directly over domain elements are present. What is missing is usually a clear process able to select and justify the selection process, that results too unsystematic and subjective.

There could different cases: elements are selected to feed an underlying computational trust model or elements may be recognized to have a meaning for trust even *without* an underlying trust model. The conceptual difference is about how much the model underlying the evidence selection is explicitly defined and how much the computation is detailed.

In the first case, the computational model defines the input that has to be collected in the application. For example, in the computation performed by Longo [Lon07], the evidence for trust is the temporal distribution of the interactions that are processed by an explicit computation that generates trust values.

A more common situation is the one in which the model of trust is rich in the human notion of trust but that loosely defines its computations.

In this case a trust model only *offers a pointer* for evidence selection and subsequent computation. Models with rich notion of trust, including cognitive models, define many ingredients and important factors to be used in a trust computation, but usually they are a high-level conceptual description with no clear computational model. They represent therefore *pointer* to a computation that defines a high-level class of evidence whose mapping to application requires work. The point is that many rich trust models do not consider the problem of evidence selection the core issue, which remains the definition of an analytical model of trust as we will discuss in the next session.

In the second case, evidence is selected without an explicit model of trust. Some heuristics or ad-hoc computation can be defined directly over some domain elements that are chosen as evidenced of trust presence. Here the starting trust model is not declared and therefore not used in the selection process.

Therefore, selection of evidence must be justified, since the element chosen was not introduced in the application for an explicit role linked to trust, but as part of the normal behavior

of the application, and no explicit model justifies it.

An alternative is to delegate the process of evidence selection to domain experts, that map trust to domain elements. This solutions has several issues: the role of trust as a distinct expertise is lost, the process is driven by domain-specific expertise and not by the expertise of trust, an expert may not be available, not accessible, and in general justifications for the conclusion could be complex, subjective or even confidential.

Conclusions

In this chapter we introduced a comprehensive notion of Trust, as it emerges from the work of Romano and the way computer scientists have translated it into a computational and formal model. The definition stresses the importance of not confounding trust with one of its components, leading to a limited and partial vision, and tries to unify all the previous trust's definitions.

Trust is seen by Romano as a complex assessment of the significance and quality of trustee's impact over trustier's expectations and sense of control in a specific situation. We concluded how this definition is compatible with the idea that trust shoul be computed by exploiting multiple mechanisms – since none of them captures trust – in a form of reasoning, the computational counter-part of Romano's assessments.

In the second part of the chapter we discussed how the notion of trust has been formalized and represented in computational models. We descussed the properties that the community of computer science's practitioners assign to the trust function, concluding how the fuzzyness of trust is reflected by its properties that are not well-defined but still identificable. We believe that this fuzzyness calls for a process embedded into the computation that checks the sustainability of specific properties of trust. This mechanism is the defeasible and presumptive reasoning, described in the next chapter 3.

Finally, in the last sections 4 and 5 we introduced our state-of-the-art review of the current landscape of computational trust models and we investigated the crucial problem of evidence selections.

We presented various mechanisms to compute trust. The majority of systems deals with scenarios based on recommadations or past outcomes processing. Computations such as the one based on similarity and categorization, the monitor-based approach by Carter or some ad-hoc heuristics

give suggestions for alternative computations. Other classical trust computation are the ones based on transitivity anf social-newtork. Promising and meaningful trust models, like the cognitive ones, has not many computational implementation we could reuse.

Regarding *conflicts resolution* among different evidence/opinions, we reported about some techniques used mainly to aggregate different recommendation values. The strategy followed is to average different values, merging different opinions and blending contradictions. In all our discussion we show the plausbile nature of a trust computation, independently form the mechanism used. We underlined how not considering uncertainty and plausibility as an active element of the computation limits many solutions. Uncertainty is usually treated with a probabilistic approach that is effective in representing and propagating uncertainy (Josang, Dempster-Shafer) but does not make explicit the reasons why a computational mechanism is more or less uncertain.

We concluded how the evidence used in trust-based computation are diverse. Anyway, if we do not consider the solutions that rely on the evalution of interactions'outcome, the set of evidence is highly reduced. We reported about several rich model of trust that define large set of evidence, but they mainly define high-level conceptual *pointers* to evidence without producing a detailed computational procedure. This is due to the fact that their main concern is the formal definition of trust rather than defining a detailed and computable evidence selection strategy.

A methodology for evidence selection that exibiths a degree of generality, that relies on the *generic expertise of trust* and able to drive the process of evidence selection over domain element does not exist. We agree with Segneur that in [Sei06] wrote how: "there is the need for a clear process between trust models and trust evidences and there are a number of types of trust evidence that have not been considered in computational trust". We showed that evidence selection becomes a complex problem when evidence has to be collected directly from domain elements and a justification is required. Anyway, this kind of selection is needed to fully reason about trust as our model aims to. Evidence selection is a key process to construct the inital arguments of our trust-based reasoning, and unexploerd evidence can extend the quality and quantity of our arguments.

In a nutshell, at the end of our state-of-the-art review, our model's design can rely on:

1. A set of well-known computational mechanisms to be translated into our defeasible trut schemes, along with with many information about how to assess their plausiblity;

- 2. A set of heuristics and few monitoring-based approaches to be made systematic;
- 3. Insights from social scientists and cognitive models about components of trust of potential computational interest.

But there is room for clear contribution in the following areas:

- 1. An efficent conflict-resolution strategy
- 2. A detailed evidence seletion strategy providing more than pointers to evidence
- 3. A computational version of many trust cognitive ingredients able to extend the trust computation in our envisaged way

Chapter 3 Presumptive and Defeasible Reasoning

Introduction

The aim of this chapter is to provide an adequate description of *Presumptive and Defeasible reasoning*, theories that stand at the core of this work. The two theories are part of an active filed in AI, the study of non-monotonic reasoning and Argumentation.

Our key hypotehsis is that trust is based on assumption and presumptions. Therefore, we need to study the structure of this presumptions, find a method to evalute the plausiblity of assumptionons on which the computation is based, and understand how to combine and aggregate the evidence in the correct way.

The theory of presumptive reasoning suggests how to define the structure and study the plausibility of each evidence. It is a form of non-monotonic reasoning based on arguments that are not completely valid, but not completely fallacious, a reasoning that increases its validity by investigating the plasubility of each argument. Presumptive reasoning checks the plausibility of each argument by using a set of critical questions fully defined by the structure of each argument, questions that test argument's background assumptions in the specific context.

A semantic of *defeasible reasoning* suggests how to combine different defeasible evidence: the validity of a reasoning based on retractable presumptions increases if the mutual relationships and the logical consistencies of the argumetns on which the conclusion is based are considered. A pure *averaging* aggregation strategy that does not consider this relationships is justificable only under complete ignorance, where infomation about the relationships among different arguments and their relative strengh is known. It represents therefore a blind aggregation, that rarely is the optimal solution. When extra information is available – or retrieved, acquired – defeasible reasoning should be used to understand better the validity of arguments: are there contradictions?, Is one argument attacking the other or baking it? A Defeasible reasoning semantic helps in answering the above questions.

Therefore, *Presumptive Reasoning* gives us a clue about how to model the content of non-monotonic argument and test it, while *Defeasible reasoning* gives us the formal tool to combine them and quantify the strength of conclusions. Walton [Wal96] suggest us the content of the arguments, while Pollock [Pol01] provide us the formal tool to combine them. They represent two complementary tools, one formal and the other descriptive, both essential in the definition of our trust models in chapter III.

The chapter is organized as follows.

In section I we introduce the basic concept of non-monotonic logic.

Then, we wonder what can be said about the structure of non-monotonic arguments, if they can be generalized, if they follow recurrent patterns and how these pattern looks like.

This is done in section 2 that introduces the theory of Presumptive reasoning by Walton, whose main contribution is the identification of a list of non-monotonic arguments – called *presumptions*- that humans use in their reasoning.

Then, in section 2 we face the problem of how to combine non-monotonic arguments by tacking advantage of their mutual relationships; how to assign different degree of justification to the arguments; how to infer conclusions from premises. We present a formal semantic of non-monotonic reasoning that differs substantially from classical deductive reasoning. This is presented in section 3, where we refer to the theory of Defeasible reasoning by Pollock [Pol1], providing justification for the selection of this specific theory. Finally, section 4 provides a state-of-the-art review about the application of non-monotonic reasoning to trust computation.

3.1 Argumentation Theory and Non-monotonic Reasoning

In this section we introduce the basic elements of Argumentation Theory and the key concept of non-monotonic logic. Argumentation Theory studies how people reason and express their arguments. It is a multi-disciplinary approach between Philosophy, Psychology and Sociology that has recently received an increasing attention from Artificial Intelligence. Argumentation is the systematic study of how arguments can be built, sustained or discarded in a discussion and the validity of the conclusions reached.

Its aim is to capture the way humans reason; it deals with the definition of the generic schemes that are used to build arguments, their validity, the dynamic of a discussion, they way they can be negotiated, retracted, used to persuade about a conclusion. Starting from the beginning of the eighties, argumentation theory has gained importance in Computer Science, with the introduction of formal and computable models of human-like reasoning. These models extended the classical reasoning models based on deductive logic that appeared increasingly inadequate for many knowledge representation problems.

In particular, the study of *non-monotonic reasoning*, commonly used by humans, has generated several new formal models of logic, such as default logic, fixed-point logic, explanatory reasoning, that have been successful used in IT applications such as knowledge-based systems, theory of beliefs, automatic proof. In the following of this section we introduce the formal notion of non-monotonic reasoning and the required terminology used in the remaining of this work.

3.1.1 Non-monotonic reasoning

Non-monotonic reasoning differs from standard deductive reasoning in the following way. In deductive reasoning, a conclusions follows form a set of true premises.

For instance, if we know that all men are mortal and Socrates is a man, therefore, following a syllogistic reasoning, Socrates is mortal. If A is the set of premise and p is the conclusion, the deductive reasoning is so that:

if
$$A \vdash p$$
, then $A, B \vdash p$ (3.1)

This means that, whatever set of evidence B we add to the set of evidence A, conclusions p is still valid. For instance, knowing something more about Socrates, true or false – such as he watches the Olympic Games – does not change the conclusion p. Therefore, deductive reasoning is

monotonic: conclusions do not change if new evidence is added in the premises, since the validity of the conclusions is all contained in the premises.

On the contrary, non-monotonic reasoning does not share property 3.1. Non monotonic conclusions can be retracted in the light of new evidence.

As an example, taken from [Bre97], let's consider the following. Bob gets up in the morning and observes that the grass is wet. He plausibly concludes that it rained overnight. When he exits home, he observes that the road is dry. He conclude that his first explanation is no longer plausible – not impossible thou, but what could have dried the road? – and a more plausible explanation for the wet grass is that the sprinkle was on last evening.

The example presented has many interesting points. It can be treated as an example of explanatory reasoning that we now formalize. Explanatory reasoning is the form of non monotonic logic that, with default logic and preferred logic represents the most prominent non-monotonic logic. It can be proven that they are equivalent formalisms that stress different aspects of the problem. While default logic and preferred logic is more a representation techniques, explanatory reasoning focuses mainly on the problem of reasoning. [Bre97]

Since we refer to these theories in the remaining of this work, we now introduce in a more formal ways these three concepts.

3.1.2 Explanatory Reasoning and Abduction

Let's T be a background theory, composed by set of logical assertions. Let's O be an observation, such as an evidence or a fact. We define *abduction* the following construct:

if
$$O$$
 and $T \vdash O$, then T (3.2)

This means that if a fact O is observed, and we know a background theory T that explains O, therefore T is a valid explanation of O. An example clarifies the definition. Suppose that our observation O is the following:

O: patient has high fever

And our known background theory T is: patients that have flu have high fever. Therefore we conclude that the patient has flu.

Abduction [Pie58] is a form of non-monotonic reasoning. If, for instance, we know that patient shows red spot on his skin, the background theory T is no more a valid explanation, or at least a weaker one.

Note how abduction reverts the classical modus-ponens construct of deductive:

if
$$0$$
 and $0 \vdash T$, then T (3.3)

Using the concept of abduction we define the concept of abductive explanation:

Let O be a set of observations and T a background theory and H a set of allowable hypothesis, i.e. the set containing all the possible hypotheses that can be formulated and proven. A set of proposition E is an abductive explanation of a proposition O iff it satisfies the following three criteria:

$$E \subseteq H \tag{3.4}$$

$$T \cup E \vdash O \tag{3.5}$$

$$T \cup E$$
 is consistent (3.6)

In the example of Bob and the wet grass, we had the following elements:

A set of proposition

$$H_1$$
: it rains

 H_2 : sprinkler is on

 O_1 : grass is wet

 O_2 : the sun shines

 $H = \{h_1, h_2\}$
 $O = \{o_1, o_2\}$

The background theory is

$$T = \begin{cases} h_1 \vdash o_1 \\ h_2 \vdash o_1 \\ not(o_2 \land h_1) \end{cases}$$
 (3.8)

And there are three possible explanations for observing O₁:

$$\begin{split} E_1 &= \{h_1\} \cup T \; \vDash O_1, and \; h_1 and \; T \; are \; consistent \\ E_2 &= \{h_2\} \cup T \; \vDash O_1, and \; h_2 \; and \; T \; are \; consistent \\ E_3 &= \{h_1, h_2\} \cup T \; \vDash O_1, and \; h_1, h_2 and \; T \; are \; consistent \end{split}$$

If O_2 is added to set of observations, the first and the second explanation are no longer valid. The goal of explanatory reasoning is therefore to find explanation that do not contains contradictions, and that justifies observations.

3.1.3 Default Logic and Closed World Assumption

Default logic [Rei80] assignes to a proposition a preferred interpretation, called its *default* true value. In deductive logic, a proposition can be either true or false.

This means that a proposition can change its default value according to new evidence, but it has a basic preferred interpretation.

More formally, a default theory is a pair (D,W). W is a set of logical formulae, called the background theory, that formalize facts that are known for sure. D is a set of default rules, each one being of the form:

$$\frac{A: B_1, B_2, \dots B_n}{C} \tag{3.9}$$

Where A, B_i and C are classical formulas. The default has this interpretation: if A (the prerequisite) is provable, and $\neg B_i$ (the consistency conditions) is not provable for every i, then we derive C. In other words, if we believe that prerequisite is true, and each of the consistency conditions are consistent with our current beliefs, we are led to believe that the conclusion is true. A classical example is the following:

$$\frac{Bird(Tweety):Flies(Tweety)}{Flies(Tweety)}$$
 (3.10)

If we know that *Tweety* is a bird, and it cannot be proven that it cannot fly, therefore it is assumed that it flies. Here the preferred interpretation is that all bird flies.

The logical formulae in W and all formulae in a default were originally assumed to be first-order logic formulae, but they can potentially be formulae in an arbitrary formal logic.

3.1.4 Closed World Assumption

As an example of preferred logic we define the closed world assumption CWA. A common default assumption is that what is not known to be true is believed to be false.

The closed World Assumption can be described with the following example. Bob goes to the station and look at the timetable. Since there is no train form Dublin to Cork on it, it concludes that there are no trains between the two cities and therefore he takes a bus. Anyway, there could be another train service that is not included in the timetable for any reason (timetable is out of date, the train belongs to another operator..). This form of reasoning is knows as argument form ignorance, since conclusion is achieved by postulating that what is ignored cannot be true. Of course, the timetable contains only positive assertions, and it does not –and it would be

impossible - to state explicitly all the negative assertions (there is no train between X and Y, X and Z ...).

The Closed World Assumption [Rei80] is a way to represent that what it is known about a problem is actually all that is known. Under this assumption the argument from ignorance is deductively valid. Formally, given a set of proposition DB (the database), the Closed World Assumption of DB is

$$CWA(DB) = DB \cup \{not(P(t))|DB \not\models P(T)\}$$
 (3.11)

That means that a proposition P(t) is added to the closed-world assumption if it cannot be inferred from the set of proposition contained in DB.

3.2 Presumptive Reasoning

Presumptive Reasoning is a type of non-monotonic reasoning where arguments and conclusions are made of presumptions. In this section we describe the theory of Presumptive Reasoning as described by Douglas Walton in [Wal96]. The theory provides us a framework to be used in our trust model and an underlying theory to justify the existence of out trust schemes described starting from next chapter.

It is important to stress the nature of Walton's work that emerges from the study of the humans' way of reasoning in a philosophical and sociological prospective. Walton's work is not a formal theory of non-monotonic reasoning; it is rather a valuable descriptive study and classifications of humans' generic presumptions.

The essential contribution of Walton is the definition of the most common presumption that humans use in their discussions. Walton's work is centred over the notion of presumptive argumentation schema, generic patterns that are used to sustain an argument in a particular context. Walton identifies 28 argumentation schema, and for each of them he investigates their plausibility with the aim not to prove their potential fallacies, but to understand their plausibility. In order to do so, he proposes and defines a set of critical questions attached to each argumentation scheme that tests its plausibility in the context.

Rather than a mere formalism, Walton's main contribution is therefore the definition of argumentation schemes and the notion of critical tests to check their plausibility.

3.2.1 What is presumptive reasoning?

Presumptive reasoning is a form of reasoning made of presumptions. A presumption is abduction: we make conclusions on the base of an interpretation –the presumption - that seems adequate to the context. In spite of new evidence, this presumption may become more plausible or so implausible that it has to be retracted.

Presumptions, observes Walton, are not deductively valid and they can be fallacious. Are they therefore useless? Or can we still use them and take advantage of their conclusions?

Walton's answer is the second. Humans use presumption in every-day discussions, often to achieve valid and useful conclusions. Even if presumption can be shown to be fallacious, they can be very effective in particular context, where they are appropriate sustainable arguments.

The key point is therefore not if a presumption is valid, but when it can be considered plausible or not. The key idea we derived from Walton's work shows how a set of tested presumptions is as effective as more complex models that claims to be closer to deductively validity.

Presumptive reasoning, continues Walton, is useful and sometime the only kind of possible reasoning. It represents a guide to prudent action trough uncertainty. [Wal96]

Presumptions are inferences taken under lack or imperfect knowledge of the context and the situation. Their validity and correctness depends on the context of dialogue appropriate for that case.

The two key concepts of Walton Presumptive Reasoning framework are the concept of argumentation schemes and critical questions.

Presumptions are rarely ad-hoc construct that are used in a dialogue. More often, presumptions are instances of generic patterns of reasoning that Walton defines as the glue that holds argumentation together in a critical discussion and makes it reasonable.

These generic patterns are known in Argumentation Theory as argumentation scheme. The existence of such schemes has always been postulated in the study of argumentation, as they appeared for the first time in Aristotle's' *Topica* (V sec. BC) [Ari28], where the author identified some recurrent types of reasoning to be used to sustain an opinion.

In order to analyse the structure of a presumption, let's consider the following example:

Bob: "He is a good baker, in 30 years I've never seen it without customers" (a)

This assertion has been produced by applying a generic argumentation pattern, named by Walton as argument form popularity to the entity *baker*. The fact that makes possible the application of

the scheme is the number of customers over time. If we know that the number of customer is actually high and constant, the presumption can be used in a discussion.

Nevertheless, Alice can reply to Bob using a critical question that attack the presumption:

Alice "Yes, because is the only one in the range of 20km!"

Alice attacks Bob argument by observing that the argument from popularity is stronger when it is actually possible to make a choice, otherwise the popularity is not an effect of the quality of the baker, as Bob thinks, but an inevitable consequence of the lack of competitions and alternatives.

The above attack to the presumption is an example of critical question attached to an argumentation schemes. The question in the example is the following: is the entity X in a dominant position? Does it have competitors?

Answering these questions changes the plausibility of the presumptions.

The central point in Walton's theory is that the critical questions are connected to each argumentation scheme and defined by the underlying theory on which the scheme is grounded. Critical questions challenges or tests the assumptions on which a scheme bases its plausibility. Therefore, critical questions are inherent to the argumentation scheme around which the presumption is build. The structure of the argumentation scheme gives a clue about how to define effective critical questions as the ones that challenges the validity of the assumptions. This observation makes the application of scheme and critical questions a kind of self-contained analysis, with implications in the way we implement this framework in a computational model of trust.

Both critical questions and argumentation scheme has to be matched to some evidence/fact of the domain in order to be applicable. This gives to a computational model of presumptive reasoning a pattern matching nature.

Finally it must be underlined the scenario in which Walton studies presumptive reasoning. Presumptive reasoning happens in a dialectical discussion, in which a *proponent* and a *respondent* are trying to support their conclusions about a topic in a context. The dialectical nature of the environment is essential to understand the nature of presumptive reasoning, as shown in the next section.

3.2.2 The dynamic of Presumptive Reasoning and the Burden of Proof

Presumptive reasoning is carried out as a dialectical argumentation between two parties (the proponent and the respondent). These two parties are trying to sustain their conclusions in a specific topic and context. Argumentation is carried on by switching the burden of proof. Every time an assertion is used in the discussion, the burden of proof is switched to the other party that has now to replicate by attacking other party conclusions or by proposing a counter-argument that contradicts them. Therefore, making a conclusion using a presumption has the effect to switch the burden of proof rather than defines conclusions. The other party will have to replicate, accepting the conclusions or trying to reject them with a counter-argument. This dynamics keeps the reasoning alive and makes the topic of the discussion clearer and the conclusions more consistent. Therefore, argues Walton, even if a presumption does not have a great plausibility in a context, it is a useful way of boosting a dialectical process by switching the burden of proof to the other party. Presumptive reasoning is therefore more effective when a dialectical process occurs.

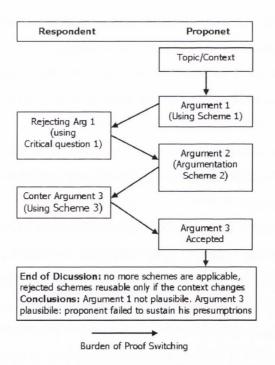


Figure 3.1 Presumptive reasoning to sustain the plausibility of a topic in a specific context. The burden of proof is switched between the proponent and the respondent.

The pattern of the discussion is depicted in figure 3.1 and it can be detailed as follows:

- 1. selection of the appropriate argumentation scheme to better fit the specific situation,
- 2. the testing of the argument answering the critical questions,
- 3. if the argument fails (i.e. other arguments or the context make it impossible to satisfy the critical questions), another scheme should be applied. If no other scheme can be applied to the specific situation, the presumption is not plausible.
- 4. If the context changes during the argumentation, arguments accepted can be rejected and rejected arguments can be re-used (i.e. presumptive reasoning is not monotonic)
- 5. When the discussion stops, arguments that could not be attacked are the valid conclusions

3.2.3 The presumptive Argumentation Scheme and the critical Questions

As described in the previous section, argumentation schemes are generic reasons to sustain an argument in a discussion.

In general, an argumentation scheme Arg_i can be represented as a set of proposition P_i that sustains a conclusion Ci. When some propositions of Arg_i are instantiated with fact/elements of the domain X, the argumentation scheme generates a presumption $Arg_i(X)$.

Each argumentation scheme has a set of critical questions Q_i representing assertions that are in accordance or opposite to the assertion contained in Arg_i .

Referring to the formal definition of abductive explanation, a presumption P is a background theory that is an abductive explanation for a fact F that is collected among the available evidence in the context of the argumentation. A critical questions check if the assertions contained in the background theory P can be attacked or backed, in order to change the validity of the conclusions. A formalization of argumentation scheme in the context of trust – our trust scheme – is provided in chapter III after the Pollock's theory of defeasible has been introduced.

For instance, let's consider the example (a) of the butcher. The used argumentation scheme, the argument from popularity, is composed by the following proposition P: "If the large majority prefers X (or thinks X is true), then there is a defeasible presumption in favour of X".

The fact used to build the presumption was:

F: "The backer's shop is full of customer".

Therefore the conclusion C: There is a presumption in favour of the Butcher's shop.

We note how the favourable presumption is better detailed by the context. In the case of the butcher, the favourable presumption could be that the shop has good product that attracts customers, for quality or price. How the presumption is used depends on the context. In our model, the favourable presumption is the trustworthiness of the entity under analysis.

The critical question used in the example is the following:

 Q_1 : There are no alternatives to X in the environment

And clearly Q_1 contrast P. Note how it does not contrast the conclusion contained in P – the baker's shop is good – but it undermines the use of the amount of customer as evidence to deduct it. AS it will be clear in the next section, Q_i is an *undercutting* defeater of P.

 Q_I is not the only critical question possible, but any reasons that undermine P is a valid candidate. Since P states that the majority has made a preference for an entity, critical questions should focus on the reasons that can explain the choice of X rather than the one contained in P's conclusions. For instance another critical question could be:

 Q_2 : Entities are not free to decide their preferences

A presumption can be used both to sustain a conclusion or its opposite, according to the fact available in the context. As argumentation schemes are not definitive evidence in favour of an argument, the same holds for the opposite.

For example, if we notice that the butcher's shop has only 10 customers per day, we might argue that this is evidence against the butcher, using the same argument in the opposite direction.

Again, critical questions are attached to the scheme, which now challenges the validity of the scheme as negative evidence. The effect of critical questions remains the same, i.e. checking the plausibility of the conclusions, but this time with the following interpretation: we tests if there are reasons that can mitigate or justify the conclusions against the object of the presumptions. In other words, we test if the negative conclusions can be reverted or justified, so that it cannot be used against the entity.

For instance, if we consider the critical question

Q3: Entity X has a limited capacity

And we observe that the butcher's shop produces exactly ten products per day, the presumption is no longer a sustainable negative argument.

We conclude by presenting some of the 28 argumentation schemes identified by Walton, including some with relevancy for the remaining of this thesis.

Argument from Expert's opinion

In this argumentation schema, the fact that E is an expert in a domain implies a defeasible reason in favour of its assertions. The schema is the following:

- E is an expert in the domain D
- E asserts that A is known to be true
- A is within D
- Therefore, A may plausibly be taken to be true

Critical questions attached to this trust scheme are:

- *Is E an expert on the field D?*
- Did E really assert A?
- *Is A really relevant to domain D?*
- Is A consistent with other experts' opinion?
- *Is A consistent with other known evidence in D?*

Argument from example

In this schema, an example E is used to sustain more generalized statement A. The critical questions are:

- Is the proposition presented by the example in fact true?
- Does the example support the general claim it is supposed to be an instance of?
- Is the example typical of the kinds of cases that the generalization ranges over?
- Were there special circumstances present in the example that would impair its generalization?

Argument from Evidence to hypothesis

This is Walton's version of abduction. If a hypothesis A is true, then B will be observed to be true.

Since B has been observed to be true, therefore, A is true. For instance: if a patient has flu, therefore it has fever. So, because the patient has fever, he has flu. Critical questions are:

- *Is it the case that if A is true than B is true?*
- *Has B been observed to be true (false)?*

• Could there be some reason why B is true, other than its being because of A being true?

Argument from Temporal projection

In this argumentation schema, if X is true at time t then it will be true at time $t+\Delta T$. This argument was studied by Pollock as well in [Pol94]. Critical questions are:

- How fast X changes?
- How big is ΔT ?
- *Was X really true at time t?*

Other presumptive scheme includes the *argument from ignorance* (whose critical question is the plausibility of the closed world assumption) and *argument from popularity* (already analysed), *argument from sign*, *argument from correlation to cause*, *argument from analogy*, the ethoic argument, the *argument from bias*.

3.4 Defeasible Reasoning and the problem of combining arguments

In the previous section we have described example of non-monotonic arguments, analysing Walton's presumptive argumentations scheme. The problem described in this section, equally at the core of our research question, is collocated in the space of formal logical calculus, and it is therefore complementary to the previous work. Walton suggests us the content of the arguments, here we investigate how to combine them in order to infer conclusions.

Reasoning is usually performed over a set of evidence, in our case defeasible arguments, from which a set of conclusions have to be inferred. This operation implies several issues:

- 1. How can we combine different arguments?
- 2. What is the nature of conclusions of a non-monotonic reasoning?
- 3. How do we assign variable degree of strength to conclusions? How do we manage contradictions?

The problem is therefore how to combine different arguments and define a way of computing the "on sum" degree of conclusions inferred. We need a semantic for non-monotonic reasoning that assigns to any set of starting argument a meaning.

The problem has been studied by several authors in the last 20 years, especially in AI context, where non-monotonic reasoning has acquired momentum due its ability to capture humans' way

of thinking under uncertainty. Here we adopt the work of J. Pollock and his theory of *Defeasible Reasoning with various degrees of justifications* as presented in his seminal paper [Pol01].

The term *defeasible* is again another way of referring to non-monotonic, where argument can be defeated or undefeated by new evidence or contrasting arguments.

Defeasible reasoning faces the problem of inferring conclusions by starting from a set including non-monotonic propositions. While deductive reasoning's conclusions are true or false, defeasible reasoning conclusions are consistent and justifiable in a given state of knowledge, called the *epistemological state* of the agent.

This helps us to answer the question number 2: regarding a set of conclusions C, all we can say is that they are justifiable and consistent if there are no arguments or evidence that can attack or contradict the propositions included in C at the present epistemological state.

The semantic for Defeasible Reasoning answers questions 1 and 3. *Defeasible reasoning* is carried on by considering the nature of the relationships among arguments and the strength of the proposition they embodied.

Arguments can support other arguments or defeat them, leading to a computation in which gradually arguments are defeated, contradictions arise or arguments becomes justifiable. In this dynamics process is therefore essential to take in consideration the mutual relationship among the arguments.

This *modus operandi* has an essential implementation in our model, where its application represents a novelty in contrast with what is usually in trust computation. Combining and computing the "on sum" of different quantified elements – being them arguments, evidence, different indicators about the status of an entity – is performed with a *blind aggregation-based* approach. Computational trust solutions falls in the *aggregation-based* approach is used instead of the method proposed in this thesis, we refer as the *argumentation-based* approach of defeasible reasoning. The difference is now analysed in details.

Argumentation vs. Aggregation

Given a set of 4 argument A,B,C,D, there are several infinite to combine them in order to achieve a final value. Suppose that each of the argument (or evidence) has a value in the set $\{-1,0,1\}$ meaning that the evidence is in positive, negative or neutral regarding the target of the computation.

The aggregation-based strategy computes the final on sum value by ignoring the mutual relationship that might occur among the four elements. The most simple aggregation strategy sums the value of the four arguments obtaining a final value.

This averaging procedure is essentially a blind practice that is performed (or justified) under complete ignorance of the relationship that occurs between the four arguments and their mutual strength.

The procedure is also justified when it is known that all the evidence are independent (no mutual relationship occurs) and they have the same relative strength, so that we can consider each of them isolated from the others and a simple sum is justified.

When something more is known about the four evidence, we should take advantage of this information to aggregate evidence in a more consistent and solid way.

If we know the relative strength of the evidence, for example that A and B has more importance than C and D, a linear combination – or any explicit formula - instead of a simple sum can be performed, assigning more importance to A and B than C and D. Even if more sophisticated, this is still an aggregation-bases strategy where four pieces of evidence are considered having more relative importance but still isolated or no information is known about their mutual relationships.

Moreover, in the final value information about the single starting value is lost.

By knowing the value of the evidence, we could look if there is an agreement among their value. For instance, if there are two arguments in favour and three arguments against, or just one argument against, these two situations lead to an aggregated value of -1 but in the first case there some elements of contradictions. It is important as well to register the number of evidence used to quantify the potential uncertainty of the value, in our example greater in the second case with just only evidence.

Nevertheless, we still not take advantage of the relationship that occurs among the evidence, but we reason over their isolated values.

When information about the mutual relationships is available, argumentation takes advantage of it. Keeping it simple, in this approach a relationship links an argument A to an argument B with two possible categories of semantic meanings: A support B or A defeats B. Suppose we know that among the four piece of evidence there are the following relationship:

B support A
$$(3.12)$$

(A and C) defeats D (3.13)

And suppose that A and C are positive evidence, B and D are negative.

The *blind-aggregation* strategy would lead to a final value of 0, meaning there is no evidence against or in favour of the target object. According to the argumentation rule 3.12, since *B* support *A*, but *B* is negative, *A* is no longer a justifiable positive piece of evidence. According to rule 3.13, *A* and *C* defeat *D*, but *A* is no longer a positive evidence, so *D* is not defeated. At the end of the reasoning, *D* is still a positive piece of evidence and *B* and *C* are still negative. All the arguments are now *justifiable* evidence, since no arguments defeated them and no argument supporting them have been defeated. Therefore, the final value aggregated after argumentation is -1, leading to a different decision that the previous one.

This simple example introduces the basic element of the way *defeasible reasoning* compute conclusions: the concept of defeaters, support evidence, justifiability and the recursion inherent to the computation. The problem is discussed and formalized around Pollock's theory in the next section.

3.3.1 Pollock's Defeasible Reasoning

According to Pollock [Pol94, Pol01], the basic idea is that reasoning consists of the construction of arguments, where reasons are the atomic links between arguments. Defeasibility arises from the fact that some reasons are subject to defeat.

Here we describe the latest version of Pollock's theory, where arguments have a variable degree of justifications. This version is used in our computational model to compute and represent trust values.

In Pollock, reasoning can be visualized as an inference graph. An inference graph is a direct weighted graph in which nodes represent defeasible argument and pair of nodes can be connected by a link.

There are two types of links between A and B: when A supports B (that means that A is a defeasible reason for B) the arguments are linked by a *support-link*, represented with a dashed line. When A defeats B (it a Defeasible reason for not sustain A), we use a solid link (*defeat-links*)

As an example, the argumentation rules 3.12 and 3.13 can be depicted as follows:

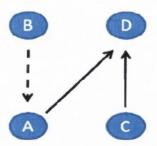


Figure 3.2 Inference graph for rule 6 and 7

We have seen that nodes can be *defeaters* or *supporters* of other nodes, or neither of the previous. Although authors do not converge on the issue, Pollock states that there are only two types of *defeaters*: *rebutting* defeaters and *undercutting* defeaters. A rebutting defeater defeats the conclusion of an inference *A*, while an undercutting defeaters defeat the link between premises and their conclusions *A*, Let's consider this example taken from Pollock [Pol01]. Suppose that an agent is watching an object *X* that appears to be red. The argument *A*: "*X appears to be red*" is a defeasible reason for assuming that B: *X is Red*. If the agent discovers that *C*: "*X is illuminated by red light*", it can no longer assume that the connection between A and B holds: *A* is no more a reason for *B*.

On the contrary, knowing that there are red lights does not defeat the fact that B:"X is red", since red objects look red as well if illuminated by red light. In other words, if A is a reason for B, after adding the undercutting defeater C, A does no longer guarantee B as before. Suppose that agent discovers D:X is blue, this is of course a rebuttal defeater, since it defeats the conclusion. Rebutting defeaters are represented by the symbol \otimes . For instance, $Arg_1 \otimes Arg_2$ means that Arg_1 no longer guarantee Arg_2 . Therefore, if C is an undercutting argument for the reason that links A to its conclusions B, we write: $C \rightarrow A \otimes B$. On an inference diagram, the example of the red light, where C is a rebutting defeater for the reason between A and B is depicted in figure 3.3.

Other authors, notably Touretzky [Tou84] identify a third type of defeater, the *specificity defeater*, which Pollock express in terns of undercutting defeaters. The idea of specificity is linked to the strength of arguments and situation where contradictions arise: if two arguments lead to a conflicting conclusions but one argument is based upon more information than the other then the more informed (or more specific) argument defeat the "less informed" one. This phenomenon is very common in reasoning and, as we will see, in trust. It is important to notice the need for various degree of strength for the argument to solve contradicting situations.

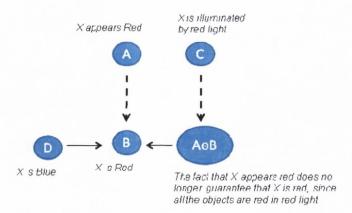


Fig 3.3 D is a rebuttal defeaters while C is an Undercutting Defeaters

Links connects root nodes to target nodes. The root of a defeater's node is a single node, while the root of a support-link can be a set of nodes. The *node-basis* of a node is the set of roots of its support-links (if it has at least one). The node-basis therefore contains all the set of nodes from which the node A is inferred in one step. If the *node-basis* of A is the empty set then node A is a *premise*.

The *node-defeaters* of a node are the roots of the defeat-links having the node as their target, i.e. all the node that potentially can defeat A. A node is *initial* iff its node-basis and node defeaters list are empty.

The status of a node can be either *defeated* or *undefeated*. Defeasible reasoning is recursive: argument A is not defeated iff its support nodes are not defeated and so forth.

Given that, we can identify three recursive rules to define how conclusions and inference are produced in defeasible reasoning:

- 1. Initial nodes are undefeated.
- 2. A node is undefeated if all the members of its node-basis are undefeated and all nodedefeaters are defeated.
- 3. A node is defeated if either some member of its node-basis is defeated or some node-defeater is undefeated.

For example, in figure 3.4 (left), E and C are initial nodes, therefore undefeated. A is undefeated since it is supported by C. F is defeated because of C, so that D is also defeated. B is undefeated since it is supported by A. The node-basis of D, for instance, is E and F, while the defeat-basis of F is C.

The undefeated arguments represent justifiable conclusions. Therefore, a *defeasible reasoning* generates a set of undefeated argument, the justifiable conclusions, and discards a set of defeated argument that cannot be justified in the present epistemological state.

The above three rules are intuitive and self-explanatory, but the recursion they contains turns out to be ungrounded. For example, considering the following graph 3.4 (right).

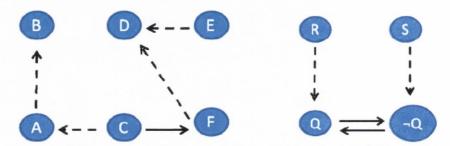


Figure 3.4 An example of inference graph (left). Q and ¬ cannot be recursively computed (right)

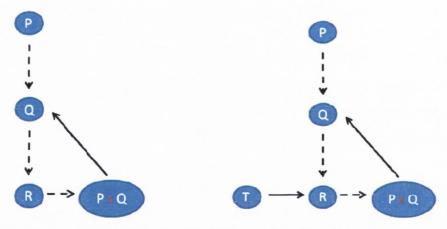


Figure 3.5 A more problematic situation (left), and the effect of an external argument T

It is not possible to assign status to node Q, since we should know the status of node $\neg Q$ that requires the status of Q. Moreover, given that P and R are true with the same degree, we have to accept that both Q and $\neg Q$ are defeated. Therefore, rules contain incorrect recursion and lead to contradictions.

The graph in figure 3.5 represents another problematic situation. If we use rules 1-3, we assign undefeated to P, and we try to assign undefeated to Q. Therefore R is undefeated, but R is a rebutting argument for Q that is no longer undefeated. If we assign defeated to Q, it means that R

is defeated and therefore Q has no defeaters. In order to fix this problem, Pollock introduces the notion of partial status assignment. The definition is the follow:

An assignment σ of *defeated* and "undefeated" to a subset of the nodes of an inference-graph is a partial status assignment iff:

- 1. σ assigns *undefeated* to any initial node;
- 2. σ assigns *undefeated* to a node α iff σ assigns *undefeated* to all the members of the node-basis of α and all node-defeaters of α are assigned *defeated*; and
- 3. σ assigns *defeated* to a node α iff either some member of the node-basis of α is assigned *defeated*, or some node-defeater of α is assigned *undefeated*.

 σ is a status assignment iff σ is a partial status assignment and σ is not properly contained in any other partial status assignment. The new semantic for Defeasible reasoning concludes that (8) a node is undefeated iff every status assignment assigns "undefeated" to it; otherwise it is defeated. Belief in P is justified for an agent iff P is encoded by an undefeated node of the inference-graph representing the agent's current epistemological state.

The idea is therefore that, given an inference graph, several consistent assignments are possible. An argument P is undefeated if it is so in all the assignments. For example, let's consider again the critical situation expressed in figure 3.4 (right) that cannot be computed with the previous rules (1-3). According to the new rule 4-6, two status assignments are possible: we assign undefeated to R and S, undefeated to Q and defeated to \neg Q. Note that this assignment is valid, Q is undefeated because the node-basis R is undefeated and the node-defeaters \neg Q is undefeated. The other assignment assigns to R and S the undefeated status, defeated to Q and undefeated to \neg Q. Therefore, R and S are undefeated (since they are so in all the status assignments) while Q and \neg Q are defeated.

Regarding the graph 3.5, the above graph is an example of collective defeat, since all status assignments lead to contradictions, therefore all the nodes are defeated according to the new semantic. Note that if an external argument T is added to the reasoning as in graph 3.5 (right), the graph is no longer an example of collective defeat, since T defeats the undercutting defeaters R of Q, that therefore is undefeated.

A further investigation of *defeasible reasoning* problem is out of the scope of this thesis and we remind to [Pol01] for further explanation. What we introduced so far is enough for the rest of our work.

3.3.1.1 Various Degree of Justifications

It is unrealistic to assume that all the arguments support their conclusions equally strongly. For instance, the arguments A_1 : "I saw the car key on the table 5 minutes ago" and A_2 : "I saw the car key on the table yesterday" have a clear different strength in supporting the Defeasible conclusions that the keys are on the table.

Given an initial or undefeated node, we call J the degree of justification (or strength) it would confer on its conclusion.

A defeasible conclusion is inferred by a premise and a reason that makes the conclusion defeasible. For instance, the conclusion that the car key are on the table are a function of the strength of the premise "Mark saw them on the table yesterday" and the strength of the reason: "if someone saw car keys on the table, they are on the table".

The strength of the premise could be connected to the fact that Mark is reliable, he is not a liar, the time he saw the keys, the fact that he does not suffer from lack of memory and his sight is good. The strength of the reason – the support link – might depend on how often things are moved from/to the table, or the probability of a statistical syllogism that says that when a person asserts X, X has a high probability of being true. We now list Pollock's generic rules for computing on sum degree of justifications.

Rule 1 - Weakest Link Principle

The first rule is the *Weakest Link Principle for Defeasible Arguments*, according to which an argument is as good as its weakest link. More precisely:

The argument strength of a defeasible argument is the minimum of the strengths of the defeasible reasons employed in it and the degrees of justification of its premises.

For instance, if it is almost certain that Mark so the car keys on the table, but the house is so messy that things are always moved apart, the conclusion will not be particularly strong. Note how, in deductive logic the consequence if "Mark saw the keys" implies that "the keys are on the table", the conclusion is fully justified by the strength of the premise, since the reasons is not defeasible. Therefore the argument strength of a deductive argument is the minimum of the degrees of justification of its premises.

Rule 2 - Accrual of reasons

The second rule applies when we have two or more independent reasons for a conclusion: does it make the conclusion more justified? Pollock define the following rule:

If we have two separate undefeated arguments for a conclusion, the degree of justification for the conclusion is the maximum of the strengths of the two arguments

Rule 3 - Influence of Defeaters

So far we did not consider the presence of defeaters. Suppose now to have the situation depicted in figure 3.6, we wonder how much the degree of justification of *S* is.

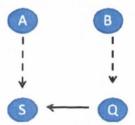


Figure 3.6 Quantifying the effect of defeaters

If the strength of Q is bigger than S, therefore S has no justifiability. On the contrary, if S is stronger than Q, the possible situations are that S can keep its full justifiability or Q has the effect of diminishing S's strength. Pollock and al. [Pol96] argue how the first situation is counterintuitive and leads to paradox (as the biased lottery paradox), and therefore the second option is the correct. This introduces the following rule: given an otherwise undefeated argument of strength x supporting P, and an otherwise undefeated argument of strength y supporting x not y and y supporting y supporting y and y supporting y supporting

(a)
$$\tau(x,0) = x$$

(b) if
$$y \ge x$$
 then $\tau(x, y) = 0$

(c) if
$$\varepsilon \ge 0$$
, $\tau(x + \varepsilon, y + \varepsilon) = \tau(z, w)$

Property (c) assumes linearity, a basic computation that Pollock proves to be compatible with the humans' intuition about reasoning. Given (a), (b), (c) the simpler admissible formula is the one that compute the strength of the resulting argument P as x-y if x>y and 0 otherwise.

Computing Defeats statuses

Since we introduced degree of justifications, each argument has now a quantitative value j attached to it rather than a binary value *defeated* or *undefeated*. Now nodes can still be defeated (j=0) or be diminished into a new value j'. For instance, in figure 3.6 Q defeats S if min(j(B),j(QB)) > min(j(A),j(AS)), otherwise Q diminishes S.

In graph of figure 3.4(*right*), representing a contradictions Pollock demonstrated (the proof is outside the scope of this work) that there are only legal computation for the degree of justification of Q and $\neg Q$ (we remind how j(R,Q) represent the degree of justification of the defeasible link from premise R to conclusion Q)

$$j(Q) = \tau \left(\min(j(R), j(R|Q)), \min(j(S), j(S \neg Q)) \right)$$

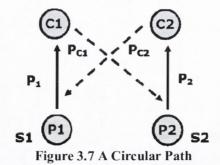
$$j(\neg Q) = \tau \left(\min(j(S), j(S \neg Q)), \min(j(R), j(R|Q)) \right)$$
(3.14)

A contradiction between two arguments is resolved by comparing strength of the arguments supporting each of them.

Circular paths

In order to conclude our description of Pollock, we describe how circular paths are treated. Circular paths occur when the status of an argument A depends on the argument itself.

An example is displayed in fig. 3.7 that is relevant to our trust-based reasoning as well. The status of C_1 requires knowing the status of P_1 that depends on C_2 that in turns is supported by P_2 whose status depends on C_1 . Circular paths are the main obstacle to compute defeasible statuses using a simple recursive procedure from conclusions to premises. Pollock treated the problem of circular path introducing the concept of equivalent graph.



101

Given an inference graph G we build the equivalent graph G^* derived from G by removing the minimum set of those links that generates mutual dependences. For instance, the equivalent graph G^* of figure 3.7 is obtained by removing the link to C_I to P_2 and the link from C_2 to P_I .

Pollock shows in [Pol01] how is legal to compute the status of C_1 and C_2 can be computed on the equivalent graph G^* and extend the result to the starting graph G.

The result obtained by Pollock is the following.

$$j(C_1) = \tau \Big(\min(j(P_1), j(P_1 C_1) \Big), \min(j(P_2), j(P_2 C_2) \Big) \Big)$$

$$j(C_2) = \tau \Big(\min(j(P_2), j(P_2 C_2) \Big), \min(j(P_1), j(P_1 C_1) \Big) \Big)$$

Therefore again the circularity is resolved by comparing the strength of the supporting argument.

3.4 Trust and Defeasible Reasoning

We conclude this chapter by investigating if defeasible reasoning techniques have been already used in trust computation. The use of argumentation and defeasible reasoning in trust, at the present time, is limited and absent in the way this work proposes it. The use of argumentation has an application in security-oriented scenarios and recently in trust management as well. These works focus on valid formal models of argumentation to be employed in a trust-based computation, usually in a multi-agent distributed scenario, but they lack *content* of trust, that is relegated to policies management. On the contrary, our aim is to provide trust with a *content*, represented by our set of trust schemes. This fact, joint to the different applicative scenarios and the different aims and tools used (mainly probabilistic tools that we avoid) make the difference with our work, despite the common idea of using an argumentative reasoning to improve the decision-making process.

We should mention three recent works (2007 and 2008), all following the first publication of our model in [Don07a].

The work of Antoniou et al [Ant07], in the context of multi-agent system modelling, wants to develop a modal semantic for distributed defeasible reasoning. Since defeasible reasoning is at the base of normative reasoning, an application to trust management, intended as distributed policies management, is envisaged. In relation to our work, their paper provides a valuable *formal* tool for argumentation, while it does not provide any content.

Kohlas et al. [Koh08] use argumentation in the context of public key authentication. In their model, the decision to trust a public key is based on binary assumptions (predicates that are assumed true or false) that are affected by uncertainty.

Using a probabilistic model, the trustier constructs all the possible scenarios, generated by combining all the possible values of all binary assumptions the trustier made. For each scenario, a probability is assigned on the base of past history information. Finally, the trustier reasons about the consistency of the conclusions of all scenarios in an argumentative way, considering conflicts and supports. The technique used is a probabilistic reasoning.

The work, even if in a different context and using different tools, results interesting for the idea of reasoning about the pieces of evidence collected by means of an argumentation approach that quantifies conflicts and supports.

Finally, the master thesis of Stranders [Str08] faces a problem similar to ours, i.e. the use of argumentation for decision-making in trust. His work has the aim to disclose the reasons behind a trust-based decision, in accordance with our thesis, but provides the formal framework for inserting arguments, not their content.

The tool used is a fuzzy argumentation, mainly centred on the use of conditional probability. Again, the work offers interesting formal methods to be employed in a trust-based argumentation (formal definition of an argument, function of preference among arguments, representation of arguments as fuzzy sets), but does not offer any innovative trust content, that is mainly relegated to subjective policies and rules. An interesting technique used is the possibilistic reasoning that starts considering the consequences of a certain decision to assess its importance during the argumentation process.

Conclusions

In this chapter we have described the fundamental elements of the theory of nonmonotonic reasoning. These elements represent the structure on which our model of trust will be defined and are at the core of our research issues.

We described the basci concept of non-monotonic logic, a form of logic where new evidence may change (defeat) the validity of conclusions. The following does not hold for a non-monotonic argument:

if
$$A \vdash p$$
, then $A, B \vdash p$ (3.1)

We introduced the basic concept of default logic, Closed World Assumption and Adbuction, all arised from the need of formalizing non-monotic logics.

Then we analysed Walton's Presumptive Reasoning and Defeasible reasoning. While the first investigate from a descriptive point of view the recurrent structure of a certain type of non-monotonic arguments humans use in their discussion, the second investigate how to formally combine them into a reasoning. We stress what we consider the main points of our analysis:

- 1. Human reasoning is defeasible rather than deductive, with conclusions that are justificable rather than valid
- 2. Even if defeasible, this form of reasoning is largely used and useful
- 3. By critically testing each assumption on which a defeasible argument is build, its plasubility increases. The tests are inherent to the structure of the argument
- 4. Defeasible rasoning combines different evidence by taking advantage of their mutual relationship and it constrasts with a simple blind-aggregation used in many computational implementation
- 5. Defeasible reasoning needs arguments with various degree of strenght

Non-monotonic reasoning is essential to model humans-based reasoning. Trust, as a human form of decision-making process, is by nature defeasible and non-monotonic and, as section 3.4 shows, the application of non-monotonic reasoning techniques to trust is at its early stage. In the next chapter, we investigate how the two theories can be used to model trust as a form of non-monotonic reasoning.

Chapter 4 A Model of Trust as Defeasible Reasoning

Introduction

In chapter 2 we introduced a definition of the human notion of trust by Romano, centred on the notion of trust as a complex evaluation of the impact of different factors linked to the trustee, trustier and context's variables. We introduced the concept of computable trust function, describing trust as an analytical function with different representation methods.

In chapter 3 we introduced the core concepts of Walton's Presumptive Reasoning and Pollock's semantic for Defeasible Reasoning, in particular the concepts of plausibility and justifiable conclusions.

In this chapter we describe how trust and defeasible reasoning can be merged together in the design of a theoretical model. We start by recalling our research issue.

Since trust is a form of defeasible and presumptive reasoning, computational techniques proper of these two fields have a positive impact on the quantitative analysis of trust.

This implies several issues:

1. The validity of the starting assumption trust is a form of defeasible reasoning

- 2. How to use defeasible reasoning and presumptive reasoning techniques in trust, i.e. how to define a computable model of trust structured over these two disciplines.
- 3. Define the positive (if any) impacts of the envisaged model in the field of computational trust: what we expect, why and how results derive form the introduction of Presumption ad Defeasible Reasoning.

Issue 3 encompasses two classes of positive impact: the theoretical contribution of the envisaged trust model and its effect over trust computation: *does it make trust computation more efficient?*Does it introduce a novel computational paradigm?

While the answer to the third issue is left to our evaluation, this chapter focuses on issues 1 and 2. In section 1 we will investigate the validity of our assumption "trust is a form of defeasible reasoning", showing how social science definitions of trust agree with this vision.

The rest of the chapter describes how we faced issue 2, which calls for the design of a trust model that accommodates presumptive and defeasible reasoning. The model design described in this chapter is kept at a high level of abstraction, while section II describes (one of the possible) computational implementations of the model.

The overall descriptive view of our model is presented in section 2, where it is made clear how the model is actually structured around the dynamics of Walton's Presumptive reasoning. In section 3 we explain the concept of trust scheme, the trust counterpart of an argumentation scheme. The complete operative list of trust schemes is left to chapter V. Section 4 describes the formal details of our models, from the representation of the trust arguments to the application of presumptive reasoning and defeasible reasoning to test the plausibility of assertion and combine them into conclusions.

In section 5-7 we formalize our theory of trust as a defeasible reasoning, defining its form (a syntax and a semantic) and its content (the generic arguments used in a trust-based reasoning). In section 5 we introduce the idea of computing trust schemes and trust values as a defeasible argumentation, in section 6 we define trust as a special inference graph, including special subsets represented by our trust scheme. In section 6 and 7 we describe the semantic used in computing our trust-based reasoning. In section 8, the design presented in this chapter let us anticipate the model's contributions and novelty brought to trust models design, that should be listed among the (theoretical) positive effects of the model.

4.1 Trust is a form of Defeasible Reasoning

The title of this section contains the starting assumption of this thesis. In line with the spirit of this thesis, we treat this not-trivial assertion as a defeasible position whose plausibility should be investigated.

We think the assertion is not trivial. If it were, trust models should contain some defeasible mechanisms and agents should reason defeasibly, ready to retract assertions and able to consider mutual relationships among pieces of evidence, fact that, to the best knowledge of the author of this thesis, is absent. The best sign of defeasibility in today's trust models is an uncertainty value attached to trust value, which implies some *caution* in using some computed trust values. Uncertainty does not capture all the reasons why arguments are defeasible. It could be the case that an argument with no uncertainty is actually easily defeated.

Let's consider the probability approach encoded in some trust value semantics, as in the assertion; trust level of Mark is 80%.

Trusting Mark is not a *sure* proposition, since 20% of the time is not true, but this is not sufficient in order to reason defeasibly. It is clear that trust is prediction, therefore subject to risk. The problem is that the mechanisms behind the computation of 20% and 80% are treated as deductively valid and *isolated*. They are given as valid starting point and not discussed, and the entire computation is based on a blind faith in these mechanisms.

For instance, since Mark did 8 correct sums and 2 wrong ones, his trust level in making sums is 80%. What if Mark had a high fever when he failed the two sums? What if the 8 correct sums were actually the same one repeated 8 times, or very easy ones? The assumption underlying the trust value is defeasible, and therefore investigating its plausibility is essential. The agent holding this information may not realize it, but another agent may attack its conclusions on the basis of this new evidence.

In order to be defeasible, a computational trust model needs a process to test the justifiability of the computation performed in the current epistemological state, and a way to combine different beliefs *defeasibly*, considering their mutual relationship, influence and consistency.

In the actual landscape of trust computational mechanism there is no sign of argumentation techniques used to solve contradictions or to combine pieces of different evidence, even if the majority of the models needs to aggregate evidence. The idea of taking advantage of the mutual relationship among evidence is neglected.

An explanation for the absence of defeasible mechanisms in trust computation could be the fact that the impact of presumption-based techniques is not considered of critical importance. This is an issue that our evaluation will directly investigate.

Given that trust is nowadays not treated as defeasible, let's go back to our starting question: is trust a form of defeasible reasoning? This implies an answer to these questions: *Is trust a form of reasoning? Is trust a form of presumptive reasoning?*

Trust as reasoning

The answer to this question emerges from the definition of trust we provided in chapter I. Romano's analysis of trust showed how trust is a complex evaluation of the impact of trustee, trustier, circumstances over a situation of interest. The concept of complex evaluation is compatible with the notion of reasoning.

Saying that trust is a form of reasoning does not seem totally easy to accept. Often humans take trust decision without reasoning, following a kind of instinct or intuition, or following unconscious actions.

According to Lagerspetz [Lag96], one loses a lot about what trust is if trust is considered solely as a reflexive phenomenon or a mental state. People in their daily lives are rarely aware of the fact that they are trusting. It is the lack of this awareness that is the expression of trust, which may be regarded as a state where we do not have particular expectations, where the situation is undefined [Lag96]. When a person starts considering his/her trust, it could be a sign of distrust.

The implicit character of tacit trust (or tacit distrust) means that possible alternatives or reasons for change are not considered. The lack of this awareness does not necessarily mean that tacitness is similar to unconsciousness. The fact that the process remains implicit does not imply that there is no reasoning in it, it might be necessary only in new or unpredicted situations. We believe that the following proposition "trust can have the form of reasoning" is certainly true, as is true that it might have other forms. Trust can be a careful evaluation of multiple evidence, some of them in favour, some against the trustee entity, reasoning performed over all the available evidence, if needed. Note that this is not a lack of trust, since the decision to be taken is still a trust-based one, and the trustee entity has the possibility to betray the trustier.

When it comes to rational agent, trust is a rational decision grounded on motivations and evidence. Saying that trust is a form of reasoning has some computational implications.

Rather than a simple formula-centred computation, saying that trust is a form or reasoning stresses the fact that a decision is taken by assembling hypotheses and beliefs rather than elaborating input data. Reasoning calls for a more logic- and explanation-based approach that is usually absent in today computational models and present in high-level models with few computational incarnations.

Trust as a presumption

Trust is a presumption for several reasons. First, trust is a decision taken in situations of imperfect-knowledge. Therefore, we perform a decision on a subset of evidence that leads to conclusions that could have been different with other evidence. Second, trust is a decision that is subject to defeat, for instance when the trustee betrays the trustier.

Accepting that trust is a risky and uncertain decision taken under imperfect knowledge is not sufficient to say that trust is a presumption. Risk or uncertainty are not presumptions. They quantify the imperfect knowledge we have, or the probability of an event to occur, but they could be a product of a procedure considered deductively valid. If we buy a lottery ticket, and we know that there is only one prize and the tickets sold are a million, the decision to buy a ticket is known to be risky and the probability of success is (deductively) 1 over a million, if we know that the lottery is not biased.

Trust, according to Sztompka [Szt00] is a bet on future contingent actions of others: i.e. one trusts that something will or will not happen in the future. This attitude is not just based on what could happen in the future, but also on a set of arguments such as past experience or recommendations. Trust is not a lottery, since it is based on some mechanisms humans have adopted to take decisions

What makes trust a presumption is the fact that the mechanisms we use to trust are mainly presumptive or defeasible. This means that conclusions strongly depend on their plausibility in the current epistemological state, and not only on the computation that the mechanism contains. The computation is therefore a presumption in favour or against trust that can be undercut or rebut.

We have already analysed one of the major mechanisms to compute trust, the past-outcomes mechanism. We can reformulate this trust mechanism in the following defeasible way:

"If entity X performed well (bad) in the past when interacting with entity Y, then there is a defeasible reason in favour (against) entity X's trustworthiness".

This is the way we believe the mechanism should be interpreted. The similarity with Walton's presumptive argumentation scheme that we shortly described is clear.

As a presumption, the mechanism has undercutting defeaters and rebuttal defeaters or, by using Walton's terminology, it has a set of critical questions attached to it. Examples could be the different difficulties of the interactions performed by X, the fact the X performed the interactions and not another entity and so forth (note how the first argument is an undercutting defeater, the second a rebutting one).

Therefore, trust mechanisms are defeasible mechanisms, assertion that is at the core of thesis. Our hypothesis is therefore that the introduction of defeasible techniques will positively impact trust computations.

The first step in order to investigate this research issue is to define how to use these techniques in trust, i.e. how trust can be modelled as defeasible form of reasoning.

Our idea is to model a trust computation around the two techniques presented in chapter II: Walton's presumptive reasoning and Pollock's defeasible reasoning semantic.

In the next section we extend this analogy that helps us to define our computational trust model.

Trust as a Presumptive and Defeasible Reasoning

There are several analogies between Trust and Presumptive reasoning that suggest the application of the second in a trust computation.

Both are practical-oriented forms of reasoning where the ultimate aim is to support a decision-making process by identifying plausible conclusions. Presumptive reasoning, wrote Walton, *is a kind of lack-of-knowledge set of inferences, a guide to prudent action through uncertainty*. [Wal96]

Walton's definition of prudent reasoning where "uncertainty prevails" fits perfectly the classic definition of computational trust by Gambetta [Gam00]: the two techniques share the basic scenario. The concept of subjective probability fits the output of a defeasible reasoning.

There are many other similarities: trust is also a non-monotonic process (new evidences or situations could drastically change the trust decision); trust is usually a dialectic negotiation between two parties (the *trustee* and the *trustier*) where one is trying to persuade the other, while the other is trying to find reasons to sustain its intuition of trusting/distrusting the trustee.

The argumentation schemes identified by Walton become trust schemes (as a topic becomes plausible because a generic argumentation scheme is applied on it, an element becomes a plausible trust evidence since a trust scheme can be applied on it). Generic trust schemes, in this case, become generic reasons to trust or distrust, reasons to be tested with critical questions exactly like Walton's argumentation scheme. A summary of these analogies is depicted in figure 4.1.

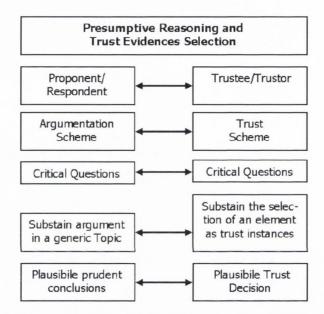


Figure 4.1 Analogy between Walton's presumptive reasoning and Trust Evidence selection

We are ready to introduce our basic assumption, grounded on this presumptive way of reasoning. A trust computation is a defeasible reasoning performed over a set of plausible trust arguments. A trust argument is composed of one of the generic trust schemes and the domain elements – the trust evidence – that support the applicability of the presumptions contained in the scheme.

The nature of this matching between application elements and trust schemes is presumptive, which means that it is not true or false in the sense of the deductive logic, but rather it is plausible or not depending on a critical analysis of its validity in that context.

Every trust scheme has a set of critical questions attached to it, a list of rebuttal or undercutting arguments defined by the structure of the trust scheme that make it more or less justifiable in the context at the present epistemological status.

Once trust arguments have been tested, a final decision has to be taken by reasoning upon the arguments, their strength, their sign (in favour or against trust) and their mutual relationships. Our idea is to introduce a Pollock-like defeasible semantic at this stage of the process, replacing the usual aggregation-based strategy.

Therefore, the trust model we present in the next section is presumptive since it treats trust mechanisms as defeasible presumptions and because it relies on defeasible reasoning semantic to combine different arguments to reach conclusions.

In order to define our trust model, and give an implementation of it, the following requirements are needed:

- 1. Define the concept of trust as a defeasible reasoning
- 2. Define a list of trust schemes, exactly as Walton defined a list of argumentation schemes, and define the critical questions attached to each trust scheme
- 3. Give a formalization of a trust scheme as defeasible argument, and of trust as defeasible reasoning
- 4. Define a semantic for trust as defeasible reasoning
- 5. Define a computational implementation of trust scheme basic computation, critical question and defeasible semantic

In the rest of this chapter we start defining our formal trust model and fundamental concepts of trust scheme required to deal with the above issues, while in the next section II we describe our implementation.

4.2 Trust as defeasible reasoning: the inference graph

Since trust is a defeasible form of reasoning, how does it look like? We represent it on an inference graph. The set of arguments of this graph supports two possible conclusions, trust or distrust. Each argument may be interrelated to the others to form a dense network. The core of this trust inference graph is represented by the trust schemes, which are special arguments that link evidence to trust, representing generic reasons to trust an entity based on a specific set of evidence.

At first level, a set of facts (available elements of the domains) are used in a defeasible computation that is attached to each trust scheme. The computation produces and quantifies evidence used to sustain the premises of a trust scheme that, joint to the presumption encoded in

it, generate conclusions. These conclusions represent reasons that are one-step from trust, and therefore we call them ingredients of trust. The premises of the trust schemes, that have to be verified in the light of the available elements, are the trust evidence.

Lagerspetz [Lag96] defined a list of ingredients of trust that we extended by merging contributions from literature, mainly Cognitive Models. As described in chapter IV, our trust ingredients are *Fulfilment, Competence, Reliability, Efficiency, Regularity, Accountability, Representativeness, Fairness, Benevolence, and Motivation*. Here we remind again how trust is for us a fuzzy but recognizable process, with specific ingredients, rules and mechanisms. Each trust scheme is a mechanism that sustains one of these ingredients.

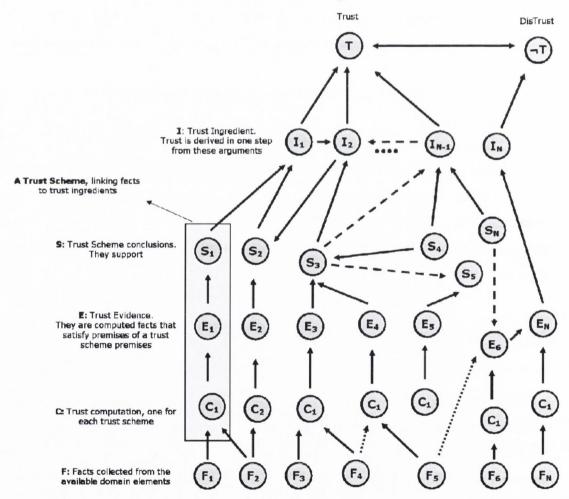


Figure 4.2 Trust reasoning: from facts to trust

In figure 4.2 we depicted our trust-based reasoning. The role of trust scheme in this reasoning is to initiate it by linking facts to trust, give a structure and a method to compute it –

collection of facts, computation of the schemes, computation of the defeasibility of each trust scheme and conclusions – and to generalize it. Note that trust schemes cannot be treated completely as isolated to each other, since it might be the case that a scheme has strong interconnections with others schemes, forming potential circular references. Due to the structure of Pollock's semantic we refer to and the structure of our trust schemes, this situation is limited and it can be simplified like this: a stage-based reasoning from facts to trust is possible without losing information.

4.2.1 Notation

We now describe the notation used in our defeasible reasoning, while its semantic is presented in 4.6. On our trust inference graph, nodes represent either reasons or assertions. A reason (defeasibly) links one or more assertions (premises) to a conclusion. An argument is generated by combining assertions and one reason.

In figure 4.3 (left) is depicted an argument whose conclusion C is derived by the premise P and the reason R according to Pollock's notation. The arc connecting the two links $P \rightarrow C$ and $R \rightarrow C$ is equivalent to a logical *and* between premise S and reason R.

In our graph, each assertion has a strength value S. Each reason-link has a plausibility value P. For the deductive reasoning-link, plausibility is set to a maximum.

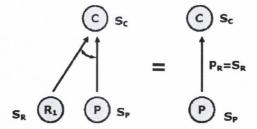


Figure 4.3 Conclusions derived from premise P with reason-link R₁

In our notation, the argument in figure 4.3 (left) is equivalent to the one in figure 4.3 (right). When a premise P is connected to a conclusion C, it means that C is inferred by P in one step using a reason whose plausibility is P_r , value associated to the link between C and P.

Another example is shown in figure 4.4. Conclusion C is inferred by starting from the same premise P using two distinct reason R_1 and R_2 . The two oriented arcs connecting R_1 and P and P and P and P show that there are two independent arguments to reach the conclusion P. The graph is

therefore equivalent to the one in figure 4.4dx that is more convenient to use during the computation (see [Pol01]).

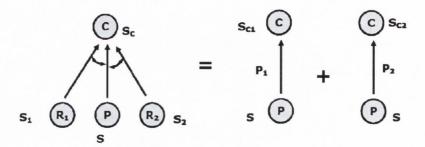


Figure 4.4 Conclusions derived from two separate reasons

Figure 4.5 (left) shows how conclusion C is derived using the reason R applied to the assertion $P = (P_1 \text{ and } P_2)$ in Pollock notation. In our notation we always depict this situation as shown in picture 4.5 (right), where P_1 and P_2 are two distinct premises joint by a deductive link to form the premises P_3 that is defeasibly linked to the conclusion C with plausibility $P = S_R$. This notation keeps P_1 and P_2 separated and it is more convenient during the computation of defeasible status of each argument when, for instance, an external argument may defeat P_1 but not P_2 and another may support P_2 but not P_1 .

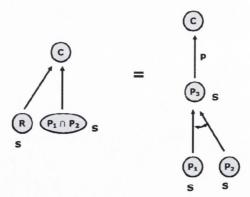


Figure 4.5 Conclusion C derived from the premise (P1 and P2)

Supporters and Defeaters

Defeaters are depicted using a dashed link. Referring to figure 4.6, a rebuttal defeater links an argument to another, such as arguments R and C. They are always symmetric, since one argument excludes the other, i.e. they represent contradictions. Rebuttal defeaters do not have any

plausibility attached to it: we will see that they are computed relying on the supporters of each of the two arguments involved in the contradiction.

Undercutting defeaters link an argument to a reason between premises and conclusion, i.e. to a reasoning link. This makes our trust graph a hypergraph. In the figure 4.6, U is the undercutting defeater of the link between P and C, influencing the value of plausibility of the reason and therefore the final strength of C. We also differentiate supporters in two groups. B is a supporter of P, in the sense that P is deducted by P and if P cannot be asserted P cannot be either. P is a supporter of the link between P and P0, meaning that it can increase the plausibility of the reason between P1 and P2, but its presence is not a necessary condition for the reason to be applied. In other words, it acts exactly like an undercutting defeater but with the opposite effect. These supporters are marked with a solid line going from the supporter to the link representing the reason

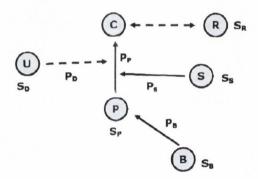


Figure 4.6 Supporters and defeaters

4.3 The (informal) notion of Trust Scheme

Trust schemes are at the core of our trust-based reasoning and its starting point. They represent generic and plausible reasons to trust an entity in a situation. Their shape and meaning is similar to Walton's presumptive argumentation scheme.

4.3.1 Trust Scheme and critical questions

A trust scheme expresses a reason why there should be a defeasible position in favor or against a trustee entity. We focus on the elements constituting a trust scheme.

The example used for our discussion is the already cited past-outcome mechanism. The trust scheme states that: if X did well in the past, this represents a defeasible reason in favor of X's trustworthiness, since we presume that X holds now the ability to deliver shown in the past.

This preposition actually hides many intermediate details of a trust scheme. Every trust scheme has premises, a defeasible reason and a conclusion. The premise of the trust scheme is "X did well in the past". This premise has to be sustained by facts collected in the domain where agents are interacting. In our model, the strength of premises is the output of a computation that is attached to each trust scheme. For instance, facts can be "Mark passed Exam 1" or "Mark failed Exam 2". The computation can be a simple percentage of the times Mark passed his exams.

The output of the computation could be "Mark passed 80% of his exams". This assertion is called *trust evidence*, since it is used to deduct the premise "Mark did well in the past". Using the reason embedded in the trust scheme, we derive conclusions from premises, in our case that "since Mark did well in the past, he still has the ability to fulfill the expectations".

The conclusion of a trust scheme does not sustain trust directly but an ingredient of trust, in our example *fulfillment*, or *ability to deliver*, that can be linked directly to trust. In general, a trust scheme has the general shape represented in figure 4.7

The way described so far is the application of the trust scheme as it was a deductive mechanism.

This way of applying a trust scheme is not defeasible, since we did not test if the reason embodied in the scheme is plausible in the context.

Defeasibility arises at each level as shown in figure 4.7. For instance, there is a defeasible reasoning link between the output of the computation and the premises: can we deduct that 80% of passed exam is an evidence that Mark did well in the past? What if the average of his class was 95%? Or 50%? These situations are strongly different.

Another defeasible reasoning link is between premises and conclusions, or between conclusions and trust ingredients and so forth. A complete example is given in section 4.3.4.

The key point is that, in order to treat the scheme as defeasible, we need to formulate questions that test the validity of the defeasible reasoning links used. These represent the critical questions attached to each trust scheme analogous to Walton's presumptive reasoning.

The critical question paradigm has two main effects:

1. To test the plausibility of the defeasible trust scheme in the context

2. To suggest to the trustier how to use the trust argument in the light of the tests' results, i.e. if it should be rejected or if maybe some of its assumptions can be re-formulated to make it plausible.

After the critical questions analysis, a trust scheme can be less or more plausible, or totally unjustified. When the plausibility is high, the trust scheme is a strong reason to trust in that context, if the evidence confirms this. P_s is not to be confounded with S_s : an argument can have a very positive (or negative) high strength of its premises, suggesting trust or distrust of an entity, but plausibility low, so that the value of S_s should be neglected and vice-versa. By considering P and S trust prediction should be of better quality.

We remind how the critical question paradigm is actually intended for dynamic discussions between two agents, one attacking or testing the arguments of the other.

In order to be critically analyzed, an agent has to declare all the facts it used in the computation, the kind of computation performed to obtain the strength of the premises, the assumption of the computation, the uncertainty of the facts used, the reason used to derive conclusions and so forth.

4.3.2 Validity of a Trust scheme

The meaning of a trust scheme has to be clarified. Trust schemes are defeasible reasons, grounded – by our assumption - on the multidisciplinary research of the human notion of trust. An untested trust scheme is a presumption, if positive it does not represents a definitive reason to trust an entity while, if negative, it is not a reason to consider the entity untrustworthy. Therefore, what validity does a trust scheme have?

It is valid in the sense that, when plausible, it is a justifiable reason to trust an entity. When it is not plausible, its usage is not justifiable. We note how critical questions have therefore the effect of reducing *false positive*, situations where implausible trust scheme are used to trust. The critical analysis of their contextual validity is more important in our method than their general validity.

Finally, a trust scheme is not to be considered in isolation. In the spirit of defeasible reasoning a trust scheme has to be evaluated in relation with other trust schemes, which may attack or support its conclusions and may increase or decrease its validity.

It should be clear that a single trust scheme is a partial view of the problem that may not be sufficient to cover the complexity of the situation. Moreover, the study of the relationships among different trust schemes may lead to more justified and solid conclusions, as our evaluation

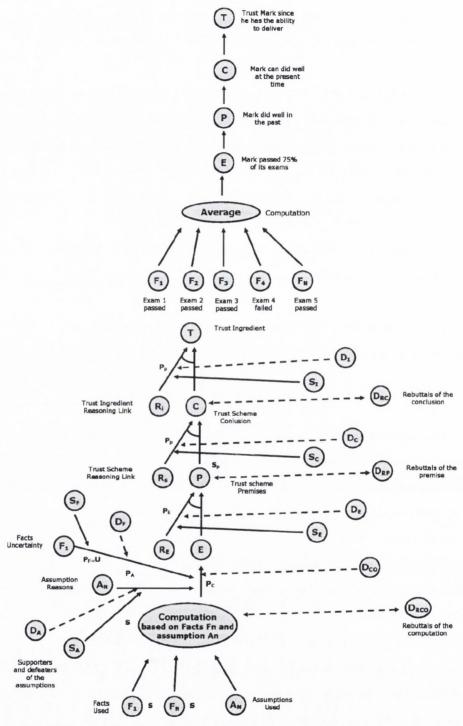


Figure 4.7 A deductive trust scheme (previous page), and a defeasible one (below). Rebuttals and undercutters are present at each level. Each reasoning-link has a base-plausibility value attached to it. Facts and assumptions are undercutters/supporters of the plausibility of the computation. chapter shows.

4.3.3 Source of Trust Scheme

It is important to underline that trust schemes are fully derived from the multidisciplinary study of trust. We strongly believe that trust schemes must be defined starting from that source, in accordance with our idea that trust is a fuzzy but defined expertise. In particular, we use social models of trust and computational models that are already part of the state-of-the-art (see chapter 2). Our trust schemes are both modeled after existing computational models – from which our contribution is providing their defeasible version and the set of critical questions – and both new computations grounded in any case in some social theory of trust, but not yet used so far.

Note that those trust schemes are domain-independent, coming from the study of trust. Nevertheless, answering some critical questions may require the investigation of domain-specific elements, even complex and requiring a specific expertise, but the process is fully defined by the structure of the trust scheme and not by a domain expert.

Moreover, the basic computation is of reasonably low complexity, since usually the reasons behind it are simple, reasonable and straightforward. This is compatible with the spirit of Presumptive Reasoning: the assumption encoded in an argumentation scheme may be usually quite simple and fragile, but it gains strength after its plausibility has been checked. Even a simple presumption, when tested, may have better results than a complicated model whose plausibility is not tested. The problem would arise if we actually considered plausible what is not.

4.3.4 Example of Trust Scheme

Figure 4.8 shows the past-outcome trust scheme on an inference graph in a potential epistemological state of the trustier agent.

In the example the trustier agent John is assessing Mark's trustworthiness relying on the past-outcomes trust scheme. The available facts are the results of four of Mark's exams. The trustier is using the percentage of good outcomes as a computation, where a good outcome is defined by passed exams (assumption A on the graph).

Facts F_1 to F_4 are certain since they are certified by Mark's school, therefore argument F supporting the validity of the computation has high strength.

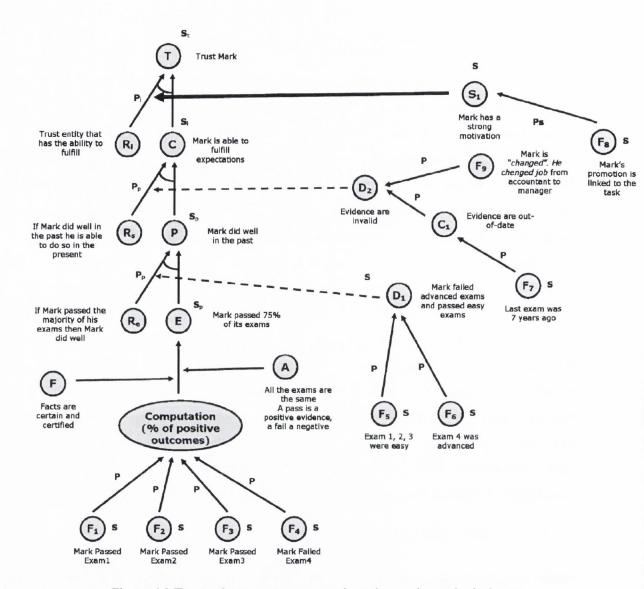


Figure 4.8 Trust scheme past-outcomes in a given epistemological state

No arguments are attacking the assumption A, while F has a high strength, so that the trustier assigns a high value of plausibility to the output of the computation, and therefore to the evidence argument E. The output of the computation is that 75% of Mark's past evidence is positive. This is used to defeasibly derive the premise of the scheme $Mark\ did\ well\ in\ the\ past$. By applying the defeasible trust scheme we derive that Mark has the ability to fulfill the expectations since he did so in the past. Ability to deliver is an ingredient of trust: the agent defeasibly deducts that he can trust Mark because he has the ability to deliver.

After a while, Bob brings to John new pieces of evidence showing that exam 4 was the most advanced and important one, while the other exams were considered quite easy. On these new facts F_5 and F_6 we derive the argument D1, stating that Mark passed easy exams and failed the important one. D1 therefore undercuts the reasoning link between evidence E and premise P, since John is now less sure that the output of the computation (75% percent of positive outcomes) is enough to conclude that Mark did well in the past (argument P).

Another fact F_7 states that Mark's last exam was 7 years ago, and F_8 says that Mark's activity is changed in the period considered, i.e. from the first exam to the present time. F_7 supports (defeasibly) the conclusion that evidence used is out-of-date. F_8 also supports the same conclusions: evidence is invalid since it is too old and entity Mark cannot be considered the same person who did Math's exams. The conclusion, argument D_2 , is an undercutting defeater of the trust scheme reason: the fact that Mark did well in the past does no longer guarantee that he still has the ability to deliver at the present time.

John also knows that Mark has a strong motivation to fulfill the expectations, since he knows that Mark's promotion is linked to the outcome of the task, so he will do his best to deliver a good outcome.

Motivation by itself represents an argument for trust, but it also represents a supporter of the defeasible reasoning going to C to T, i.e. a supporter of the trust scheme. Note how this is an example of exploiting the mutual relationship among the trust scheme: knowing that Mark has the ability to deliver and he has a strong motivation represents a much stronger argument than the argument "Mark has the ability to deliver" by itself. Moreover, we note how the supporter S may have no effect if the undercutting defeater D_I and D_2 have already invalidated the trust scheme. In future epistemological states new pieces of evidence may undercut D_I and D_2 .

Where does plausibility come from?

We now analyse the meaning of the plausibility factor attached to defeasible links. In general, the value of plausibility is an estimation of the strength of the reason with which a conclusion is deducted starting from the premises. In other words, it is an estimation of the degree of defeasibility of our reasoning. Our plausibility value encompasses different sources, each of them representing an argument defeating or supporting the reasoning-link. Sources of plausibility that we consider are the following:

Uncertainty. Uncertainty derives from the fact that data are only partially available or facts are collected using measurement tools subject to error. Usually in our computation uncertainty is present for the following reason:

- 1. Sampling and missing data. We perform a sampling over a sample of size S taken from a set of elements of size E in order to derive conclusions valid for all the elements of the set. The uncertainty can be quantified using statistics as shown in the next chapter, and regulated by enlarging or reducing the size of S. A similar situation is when partial data are available, with the only difference that we do not control which data have to be considered and which ones can be discarded. For instance, past Maths exams that Mark did in 2006 might not be available. Missing data can be interpolated using available data and conclusions are affected by an uncertainty proportional to the size of missing data.
- 2. Not verifiable data. Data used are complete, potentially not affected by errors but not verifiable; therefore their usage is based on the assumption that they have a specific meaning. For instance, a trust computation may be partially based on a user profile created by the user, where the information inserted are not easily or not at all verifiable.
- Uncertainty in the source. Uncertainty derives from the usage of unofficial sources of
 information, whose reliability is unknown, or by the fact that the techniques or procedures
 adopted to gather input data are affected by uncertainty.

Amount of knowledge. Carbone and Nielsen [Car05] already introduced a partial ordering on Trust based on the amount of evidence available for a specific entity E. If A and B have a similar trust value, but we hold more evidence about A, the trust value associated to A is more plausible and better grounded. The amount of evidence is therefore a factor that might defeat conclusions. On our inference network, the amount of knowledge represents an argument supporting or defeating (undercutting) a reasoning link.

P₀: Base value of plausibility. The term *base plausibility* refers directly to the reason that links premises to conclusions, regardless of the amount of knowledge contained in the premises and the uncertainty of data. The base plausibility is an estimation of how much the reason contained in the link between premises and conclusion is defeasible. An agent may estimate the base plausibility by considering element of the domains, historical data or past experiences,

domain-specific expertise involving situations in which the reasoning-link was involved, the structure of the computation. The base plausibility is the value of the plausibility in absence of any other external factors, i.e. defeater or supporter arguments.

The final value of plausibility is therefore the base value modified by the effect of those arguments.

The base value is a dynamic one: in the light of new evidence or interactions – not changing the potential supporters and defeaters - an agent may realize that it over/underestimated the base value for that reasoning-link and may decide to change it.

Therefore, our term *plausibility* encompasses the base plausibility and the effect of uncertainty and the quantity of knowledge available.

An example of how a group of agents might be ranked according to the level of plausibility of different computations used and the amount of knowledge available is described in appendix D.

Where is defeasibility on a trust scheme?

After having described the source of plausibility, we now accommodate all the sources of plausibility on an appropriate defeasible version of the generic trust scheme depicted in figure 4.7 above. Defeasibility arises every time the scheme inputs, assumptions or reasons are subject to defeat.

A first instance of defeasibility is represented by the computation that quantifies the trust evidence (argument E). The computation chosen is defeasible for the following reasons: it is based on assumptions that can be defeated and it requires facts that can be affected by uncertainty or that represent a partial view of the problem. Therefore the computation used in the trust scheme has always the following potential undercutting defeaters: A – the set of assumption used in the computation –, the uncertainty of F – the facts used in the computation – and K –the amount of knowledge used by the computation. These arguments modify – supporting or undercutting - the plausibility of the computation.

In figure 4.7, we depicted with S_a , D_a , S_f , D_f the undercutting links (defeaters or supporters) of assumptions A and facts F.

A second source of defeasibility is represented by all the reasoning-links going from the premise P to the conclusion T.

First, there is a reason-link that derives the premise P from the trust evidence E. For instance, from E: "Mark passed 80% of its exams" we derive the argument P: "Mark did well in the past". This deduction is clearly defeasible, since we presume that 80% of passed exams are enough to

state P. In the light of new evidence, it might be the case that 80% is no longer enough to state P (suppose that it turns out that all of Mark's fellows passed 100% of exams), or the opposite: if the average of passed exams was 50%, P can be stated with increased strength. Therefore, there is a value of plausibility P_e attached to the reason-link from E to P, affecting the strength of the premise P. On graph 3.7, this defeasible nature is expressed by the reasoning-link R_e of plausibility P_e , and the presence of defeaters D_e or supporters S_e of P_e .

Another defeasible reasoning-link is between premises and conclusion of the trust scheme (from P to C). Given that Mark did well in the past, we presume that he has the ability to deliver at the present time. Clearly this is an assumption that can be defeated. For instance, if a new evidence shows that the entity Mark changed drastically by the time he did the last exam, we are less sure about the strength of the conclusions.

The defeasible nature of the link premise-conclusion is represented in 3.7 by the reasoning-link R_p with plausibility P_p , defeaters D_p and supporters S_p .

Finally, the last source of defeasibility is the link between output of the trust scheme C and trust T. The argument C: "Mark has the ability to do well at the present time" can be clearly linked to trust, but it does not guarantee trust. If we know that Mark has a lack of motivation, even if he is able to fulfill, he might not satisfy the trustier's expectations completely. Therefore, we add another defeasible link R_i between C and T with plausibility P_i , defeaters D_i and supporters S_i .

Finally, in order to keep a link with Walton's theory, we define as *critical questions* any reasons that generate a defeater or supporter of any part of the trust scheme.

4.4 Overall design of the method, operational description

Our model's goal is to reason defeasibly about trust. The arguments of such a reasoning are identified by applying a set of generic trust scheme over the available domain elements. These arguments are then quantified and tested before being combined into a defeasible reasoning process. The final output is therefore a set of justifiable trust-related conclusions in the current epistemological state, on which a decision-making process can be performed.

Our model is composed by a set of trust schemes TS_n , each of them with its set of critical CQ_n . These two sets represent our Trust Expertise, the corpus of knowledge derived from the social and computational study of trust. They are reasons to trust an entity, and, exactly like Walton's Schemes, they have critical questions attached to them in order to make them plausible. Not only

is trust a form of reasoning, but we also provide the recurrent patterns of this reasoning. The role of the trust schemes is critical: they represent the backbone of our reasoning since they initiate the reasoning, they provide its arguments and by linking domain elements to trust they provide a strategy to identify trust evidence.

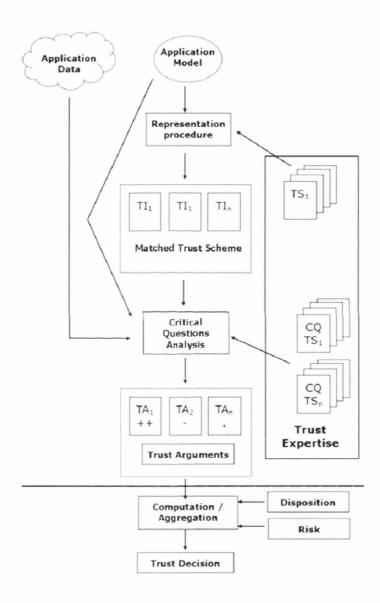


Figure 4.9 A model of Trust based on Defeasible and Presumptive Reasoning

Trust is therefore a definite object in our model and it could not be otherwise: no plausibility study or defeasible reasoning could be possible without requiring an explicit definition of the reasons why an entity should be trusted.

The model is applied in 4 stages: evidence selection, basic computation and critical question analysis, defeasible argumentation, and the trust-based decision. In the rest of the discussion, we refer to the trust scheme Past Outcome as an example.

Figure 4.10 illustrates another high-level representation of our method and its four stages.

The trust reasoning starts from facts F_n , on which a series of trust schemes T_s is matched. The same facts can be matched by more than a trust scheme. This phase is our evidence selection stage.

After the trust schemes have been matched, we test them with a set of critical questions Q_n , peculiar to each scheme, that can strengthen (solid line) or attack (dashed line) the validity of a trust scheme.

Tested trust schemes become trust arguments (represented by C in figure 4.10). Note how some trust arguments can be weaker, totally defeated or strengthened after the critical questions analysis. An argument supports trust or distrust. Arguments enter now the defeasible argumentation phase, where their mutual consistency is checked. Some arguments support others (such as C_2 and C_3) or attack them (such as C_5 with C_1 , C_2 and C_3). The final decision is taken on the basis of the arguments that survived the defeasible argumentation (and therefore are justifiable) and by comparing the strength of the argument supporting trust and distrust.

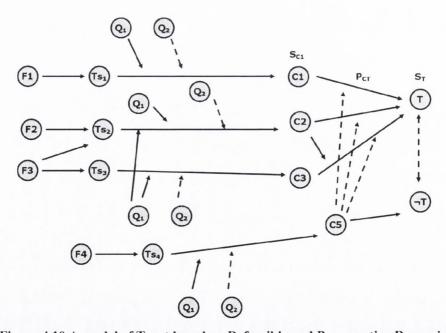


Figure 4.10 A model of Trust based on Defeasible and Presumptive Reasoning

4.4.1 Stage 1: Evidence Selection

The starting point of the process is the phase of evidence selection. By starting from a representation of the domain where the trust decision has to be taken, we need to identify which elements can be used as trust evidence. We define *trust evidence* any elements that can be used to satisfy a trust scheme's premise.

Trust schemes justify why an element should be selected as trust evidence: the element could be an input needed for a trust scheme or be a complete instantiated version of it.

We identify trust evidence by starting from a representation of the application/domain, represented in figure 4.9 as the *Application Model*.

The goal of the modeling phase is to gather all the information needed to support the presumptive identification and the critical analysis of trust evidence. Information, already present in the starting representation or to be added on the application model includes:

- analysis of entities: properties, actions, recognizability, relationships, dependency, topology, goals.
- analysis of actions: effort to complete, effect for the entities involved, dependency among actions, consequent actions, impact on entities, impact on the environment, frequency of interactions, possible outcomes of an action, observability of the outcomes.
- how communication is possible (a requisite for making indirect trust more plausible)
- analysis of the environments: dimension, how it changes in time (support persistence) and its dimension.
- memory constraints present regarding entities, actions, objects (value of memory in trust for direct experience, pluralism)

The definition of a complete trust-aware representation procedure is out of the scope of this thesis, where UML diagrams are used in all our experiments described in chapter VIII. A trust aware representation should be an augmented representation containing all the information that our trust schemes require added to a basic underlying representation.

Using the application model, the identification of trust evidences could be conveniently described as a kind of pattern matching. In order to be matched, each trust scheme has a set of conditions that domain elements have to satisfy. For instance, the past-outcome trust scheme requires that there is an action A performed by a trustee entity X, whose outcomes are observable and quantifiable by an entity Y in order to be applied.

Obviously an element or action can be instantiated by more than one trust scheme and vice-versa: the perception of the same element depends on the particular point of view and the collection of all point of views can produce a better understanding of the unbounded problem.

Once the trust evidence has been collected, then the next phase is the computation of the trust schemes and their critical analysis. Note how the applicability does not guarantee anything on the effect of the trust argument.

4.4.2 Stage 2: Trust scheme Computation and Critical Question Analysis.

In this phase, using the selected trust evidence, the computation C attached to each scheme is performed. For instance, in the first phase we identified the list of Mark's past results of Math's as trust facts; we now apply the basic computation in order to obtain a value V_s for the scheme of 0.7

Critical Question Analysis

In the Critical Question Analysis, the trust conclusions quantified in the stage 2a are tested against critical questions. These questions test the validity of the scheme applied on a particular entity in the context of interest. This stage may require information coming from the application metadata (model) and data coming from the application as well.

As described above, the critical questions undercut or support a trust scheme conclusion by modifying its plausibility value Ps. As a result of this analysis, the trustier agent may decide to:

- 1. rely on the trust scheme
- 2. reject the trust scheme
- 3. seek alternative computations or extra evidence that can neutralize and address critical questions

It might be the case that there are no alternative computations, or no extra evidence can be collected to neutralize potential defeaters or supporters of the scheme.

For instance, the past-outcome trust scheme has among its critical questions the following:

- 1. Are the outcomes collected of comparable complexity?
- 2. When were the past outcomes collected?
- 3. Has entity X changed during the time evidence was collected?

Let's suppose that Mark's ten tests were the ones showed in figure 4.11

| Mark's last 10 tests evaluation | | | | | | | | | | |
|---------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Test | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Mark | Passed | Failed | Passed | Failed | Passed | Passed | Passed | Failed | Passed | Passed |
| Complexity | Medium | Medium | High | Medium | Low | High | Medium | Medium | Medium | High |
| Date | 12/02 | 11/03 | 2/05 | 5/05 | 09/05 | 11/05 | 2/06 | 5/06 | 09/06 | 11/06 |

Figure 4.11 Evidence collected about Mark's exams

The plausibility of the basic computation is not high, as the critical question 1 underlines. This means that we cannot rely too much on the average value, or we could use an alternative computation. The critical question 2 suggests performing a computation where the older outcomes have less value, and again make the computed value weaker, since the evidence collected span almost four years. The critical question 3 has no effect.

The critical question 1 implies that the complexity of the tests is classified, maybe with the help of a Math's lecturer and it has the effect of giving more value to the 3rd, 6th and 10th tests and less to the 5th.

For instance, a trustier agent may assign, as alternative computation, a different weight according to the year of the test and its complexity, obtaining a new value V_s of 0.75 that reflects the fact that Mark always passed the complex tests and his failures are mainly in the past. The value of plausibility P_s , due to the alternative computation, is now higher.

4.4.3 Stage 3: Defeasible Argumentation (informal description)

The output of stage 2 is a set of trust arguments with a certain strength and plausibility. These sets of arguments may contradict each other, and have complex mutual relationships of

attack or support. A strategy to solve conflicts and support a decision is required. As already

anticipated, the defeasible reasoning semantic is our way to solve this problem.

The classical solution in computational trust is based on an aggregation of each piece of evidence. Many variations are possible, such as using probability distribution functions to weight contributions, but the key idea remains the same: isolated pieces of evidence are aggregated into a single value. We see clear disadvantages in a pure aggregating solution: differences are merged and blended; contradictions are lost, mutual consistency ignored. We have already described in

chapter II how a pure aggregation strategy is justifiable only under complete ignorance of the mutual relationship among the different arguments.

In this stage of the model we perform a defeasible argumentation over the trust arguments in order to identify a set of justifiable arguments.

Our trust-based defeasible reasoning has the general shape depicted in figure 4.2. At the top of the reasoning there are the two rebuttal arguments trust (T) and distrust $(\neg T)$. These two arguments are supported by a set of arguments that may defeat each other, so that at the end of the reasoning only justifiable arguments survive.

We are therefore in the same situation described in chapter II, section 2.5, where two conflicting arguments are supported by a set of distinct arguments. The conclusion, in line with Pollock, will be taken by comparing the strength of the supporting set of arguments for trust and distrust, derived by the application of trust schemes.

4.4.4 Stage 4: Exogenous factors and final decision

A final trust-decision is taken considering some exogenous factors independent from the evidence-based computation performed so far. Our model encompasses the two exogenous factors risk and disposition. A detailed analysis of their role is outside the scope of this thesis. Here we provide a possible simple model for these two concepts – well known in trust models - compatible with the overall design.

Risk could be classically modeled as a threshold, function of the trust value produced after stage 3. If stage 3 produced a trust value *T*, a trust-based decision is taken only if

$$R(R_0, T) > R_T \tag{4.1}$$

where R_T is the level of risk that the entity is willing to accept, while $R(R_0, T)$ is a function expressing the risk of a situation at a given trust value T. $R(R_0, T)$ should be a decreasing function of T: risk is mitigated when the level of trust increases, as already noticed by several authors [Cah05], [Sei06], [Mar94].

 R_{θ} is the level of risk that the trustier assigns to the situation independently from the trust level of the trustee. Note how $R(R_{\theta}, T)$ could be equal to R_{θ} , if the trustier does not consider trust able to mitigate risk. In this case the impact of the trustee is neglected and a decision is taken solely on risk: all the trust computation described so far is not necessary. Usually, given the value of R_{θ} and R_{t} , we need a certain level of trust in order to start an interaction.

Regarding disposition, it refers to a generic attitude of the trustier. The disposition of a trustee, similarly to risk, impacts the level of trust required to start an interaction. An optimistic trustier considers more the positive arguments in favour of the trustee than the negative ones, and vice versa.

How can disposition be inserted in our method computation?

We still use the classical threshold approach. The novelty is that the threshold is applied over a value coming from an argumentation of different evidence and not from an aggregation.

A second novelty is represented by the use of arguments rather than values. In our argument-based model optimistic and pessimistic attitudes can be easily modeled. A fully optimistic agent will consider only the positive justifiable arguments in order to grant trust, ignoring or minimizing the negative ones, while a pessimistic agent won't grant trust in the presence of a single piece of negative evidence against the trustee. All the other cases are in between.

For instance, after the defeasible reasoning a trustee has 2 arguments in favour and 1 against him. An optimistic trustier will produce a final trust value by considering only arguments in favour – he looks for reason to trust an entity –while a pessimistic agent will consider only the ones against the trustee – he is looking for at least one reason not to trust him. The in-between agent will aggregate the arguments with a linear combination that gives more weight to positive or negative evidence.

Note how this linear combination is an aggregation performed after an argumentation; its results are generally different from the ones obtained by a pure aggregation strategy.

The final trust decision is therefore a function of:

- a) The trust arguments/value generated after Stage 3
- b) The disposition of the trustee
- c) The level of Risk (function R)
- d) The threshold needed to start an interaction T_f

The threshold T_f is linked to the level of trust and differs from the value of R_t . The trust value could be enough to satisfy R_t but not enough to satisfy T_f , the minimum value of trust required to initiate an interaction.

If the trust value is enough to generate an acceptable risk for the trustier entity, and it is enough to initiate an interactions and it is compatible with trustier's dispositional threshold, then trust can be granted.

4.5 Formal Definitions of Trust Scheme

After having described our method stages, we provide a formal definition of trust reasoning graph and trust scheme, while in the next section we will describe the semantic for computing defeasible status of arguments.

4.5.1 Trust-based reasoning Graph

A trust-based reasoning is a hypergraph *T* defined as follows:

1.
$$T = (Arg, L_{rs}L_{u}, L_{r}, F)$$

2.
$$L = L_{rs} \cup L_u \cup L_r$$

3.
$$F_{link}: Arg \times \{Arg \cup L_r\} \rightarrow L$$

4.
$$b \in L_r$$
, $(F_{link}(a,b) = l \iff l \in L_u)$

5.
$$trust \in Arg \land \neg trust \in Arg$$

6.
$$\forall i \in I, I \subset Arg, \exists r \in L_{re} | F_{link}(i, trust) = r \lor F_{link}(i, \neg trust) = r$$

7.
$$F \subset Arg, \forall f \in F, \forall a \in \{Arg \setminus F\} | F_{link}(a, f) = \emptyset$$

8. if
$$F_{link}(a_1, a_2) = l \wedge l \in L_r \rightarrow F_{link}(a_2, a_1) = l$$

9.
$$\forall f \in F, f = (statement, u), u \in \mathbb{R}$$

10.
$$\forall l \in L_u \cup L_{r_g}, l = (statement, p), p \in \mathbb{R},$$

11.
$$\forall a \in Arg, a = (statement, s), s \in \mathbb{R}$$

12.
$$F_{agg}: \mathbb{R}^N \to \mathbb{R}$$

13.
$$F_c: \mathbb{R}^2 \to \mathbb{R}$$

14.
$$F_a : \mathbb{R}^3 \to \mathbb{R}$$

The above formulas have the following meaning. A trust graph T is a 5-tuple composed by:

- Arg: the set of all the arguments
- L_{re} : the set of all the reasoning links (defeasible and deductive)
- L_u : the set of all the undercutting links
- L_r : the set of all the rebuttal links
- F_{link} : the function of adjacency, that gives the links between two arguments or between an argument and a reasoning-link.

Definition 4 asserts that only undercutting links can connect argument to reason-link. Definition 5 states that trust and distrust are always part of the graph, while 6 defines the set *I* of

trust ingredients as all the arguments that are directly linked to the argument "trust" or "distrust". Point 7 defines the set of facts F as the leaf nodes of the graph, i.e. the premises. Definition 8 states that rebuttal links are always symmetric. Definition 9 states that a fact F is composed by a statement and an uncertainty value. Reasoning links – rebuttal or undercutting – have a statement, a value of plausibility P. Finally, definition 11 states that an argument is composed by a statement and a value of strength S.

In order to compute the strength of each argument, three functions are also introduced (definitions 12-14).

Function of Aggregation F_{agg}

 F_{agg} defines how to compute the strength of an argument supported by n different reasons.

$$F_{\alpha,q,q}:\mathbb{R}^N\to\mathbb{R}$$

 F_{agg} is attached to each argument of the trust graph except facts F.

Function of Conclusion F_c

 F_c defines the strength of a conclusion deducted from a premise of strength S and a reasoning link of plausibility P

$$F_c: \mathbb{R}^2 \to \mathbb{R}$$

All the reasoning and undercutting links (supporters or defeaters) has this function associated.

Function of Attack Fa

 F_a defines how the plausibility of a reasoning link R changes when attacked/supported by a link B of plausibility P_b connected to an argument of strength S_b . The new plausibility is a function of three values: the previous plausibility of $R(P_r)$, P_b and S_b .

$$F_a: \mathbb{R}^3 \to \mathbb{R}$$

The three above formula will be analyzed in the next section dedicated to our semantic.

4.5.2 Trust scheme definition

A trust scheme is a graph defined as follows:

1.
$$T_s = (Arg, L_{rs}, L_u, L_r, F_{link})$$

2.
$$Arg = F \cup A \cup \{c_o, e, p, c, i\}$$

3.
$$R_n, R_n, R_n, R_i \in L_n$$

4.
$$F_{link}: Arg \times \{Arg \cup L_r\} \rightarrow L_{rs} \cup L_d \cup L_r$$

5.
$$F_{link}(c_0, e) = R_e$$

6.
$$F_{link}(e, p) = R_e$$

7.
$$F_{link}(p,c) = R_s$$

8.
$$F_{link}(c,t) = R_i$$

$$9. \ \ \, \forall f \in F, F_{link}(f,c_0) = l_{r1} \Longleftrightarrow F_{link}(f,c_0) = l_{r2} \, , \\ f = (st,u), l_{r1} = (st,1), l_{r2} = (st,u)$$

10.
$$\forall a \in A, \forall x \in Arg, F_{link}(a, R_c) = l_{r1}, F_{link}(a, x) = \emptyset$$

11.
$$\forall x \in Arg, F_{link}(k, R_c) = l_{r1}, F_{link}(k, x) = \emptyset$$

A trust scheme is a hypergraph composed by the same elements as a trust graph, with some constraints on the way arguments are connected.

The set Arg of arguments of a trust scheme contains always a subset composed by:

- F: a set of facts
- A: a set of assumptions
- C_0 : the trust scheme computation
- *E:* the trust scheme evidence, that is the argument derived by the output of the computation
- P: the trust scheme premise
- C: the trust scheme conclusion
- *I*: the trust ingredient supported by the trust scheme

These definitions formalize all the components of a trust scheme such as the one depicted in figure 4.8. F_{link} is the usual function of adjacency.

The reasoning links R_c , R_e , R_p , R_i are always present and they link, respectively, computation output to evidence (Rc), evidence to premise (R_e), premise to conclusion (R_p) and conclusion to trust ingredients (R_i) (definitions 4-8).

The meaning of constraint 9 is the following: facts are connected to the computation c_{θ} with a deductive reasoning-link (L_{rl} has plausibility 1) and they are also connected to the reasoning-link R_c between evidence and computation to form an undercutting defeater of plausibility u. Definition 10 states that assumptions can only be an undercutting link of R_c .

Finally (def. 11), k represents the argument about the amount of knowledge used in the trust scheme computation. k is an undercutting defeater of R_c . The strength of k will be proportional to the amount of facts collected for performing the computation. The plausibility of the undercutting

argument k will be set to a default value P_{θ} and it will be attacked or supported as any other argument. Note how k will not be omitted in all our examples.

4.6 Semantic

The semantic of our trust-based defeasible reasoning contains the rules to associate a value to each element of the inference graph, in a way that is meaningful for the agent taking the decision and in a way that captures a compatible human-way of reasoning. We largely start from Pollock semantic, that we extend and specialize for our specific trust-case.

Rule 1: Strength of a conclusion - function Fc

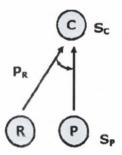


Figure 4.12 A defeasible reasoning link

The basic rule allows us to compute the strength of a conclusion given a premise P of strength S_p and a defeasible reasoning-link R of plausibility P_r .

Pollock used the weakest link principle to deal with this situation, so that

$$S_c = min(S_p, P_r) \tag{4.2}$$

If we follow Pollock's rule, the minimum of the two values will define the strength of the conclusion. Another candidate semantic suggests interpreting plausibility and strength as probabilities and, according to probabilities calculus, using the conditional probability rules to derive the strength of conclusions. This choice would lead to a strength of conclusions proportional to the product of the strength of premises and link plausibility, i.e. proportional to $S_{\nu}P_{\nu}$.

This potential semantic, known as the *probabilistic model*, is strongly criticized by Pollock et al [Pol01], who reject the assimilation of arguments' strength to probability. The most used

argument against probability is the fact that the reasoning chain, even if composed by solid arguments, rapidly decreases its strength, so that it becomes impossible to justify the deductions. For instance, as described by Pollock, the probabilistic model would make it impossible for a conclusion to be justified on the basis of a deductive argument from numerous uncertain premises. This is because as you conjoin premises, if degrees of support work like probabilities, the degree of support decreases. Suppose you have 100 independent premises, each highly probable, having, say, probability .99. According to the probability calculus, the probability of the conjunction will be only .37, so we could never be justified in using these 100 premises conjointly in drawing a conclusion. But this flies in the face of human practice. For example, an engineer building a bridge will not hesitate to make use of one hundred independent measurements to compute (deduce) the correct size for a girder.

On the other end, the principle of the weakest link is not completely out of criticism. If we consider the following two situations: (i) a premise P_1 of strength 0.5 and link R_1 of strength 0.5, (ii) a premise P_2 of strength 0.9 and link R_2 of strength 0.45, according to the weakest link principle the conclusion 1 has a strength of min(0.5,0.5)=0.5 and the conclusion 2 a strength of min(0.45,0.9)=0.45. This behaviour is not always acceptable: even if a reasoning link is weaker than another, the different strength of the premises should affect the strength of the conclusion. In other words, using the weakest link principle we lose information about the max between P_r

Since in our trust-based reasoning premises are represented by the output of a computation attached to a trust scheme, their impact on the strength of the conclusions cannot be lost. Therefore, we decide to define our semantic as follows. Given a reasoning composed by n reasoning links, the strength of the conclusion is computed as follows:

and S_p that could be important.

- a. The strength of the conclusion is computed using the weakest link principle, recursively from the final conclusion of the reasoning back to the premises, if the premises are not a premise of the entire reasoning, but an intermediate step.
- b. The strength of the conclusion is computed by the product $S_p P_r$ when the premise P is a premise of the whole reasoning, i.e. is the output of a computation based on a set of facts. In this way, the strength of the premise cannot be neglected in any case in the final conclusion, and the reasoning chain does not suffer from the probabilistic model weaknesses.

For instance, referring to the trust scheme of figure 4.8, only the conclusion P is computed with the rule b, while all the subsequent conclusions are computed with rule a.

Case of deductive links

Assuming a value of P in [0,1], when R is not defeasible – i.e. deductive – the plausibility is equal to certainty ($P_0=I$) and $S_c=S_p$. The link can still be attacked if data used to assert premises are affected by uncertainty, or if premises are asserted based on a different amount of knowledge.

Rule 2: Activation of links.

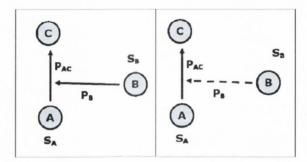


Figure 4.13 Supporters and Defeaters

If A is in relation with B, the relation is active only if the plausibility of the two trust arguments satisfy specific conditions. The idea is that a trust argument can support or defeat another one only if its value of plausibility/strength is adequate in comparison to the other argument.

Our rules are different for supportive links and attacking links. A supports B only and only if A is more plausible than B and it generates a stronger argument than B. Therefore the following must hold:

$$S_{ca} - S_{cb} \ge T = 0 \qquad (4.3)$$

where F_c is the function of conclusion. T is a threshold that we set up to zero in our discussion but that in general can have a positive value to express more relaxed activation rules.

The rule has also an intuitive interpretation in the way we reason: when an argument is attacked by another one of lower plausibility or uncertainty, it is easy to counter-attack the argument on these bases.

Regarding the role of defeaters, the situation is different. A defeater, stronger or weaker, represents an argument that contrasts the starting one. Therefore, its role cannot be neglected in any case. Of course, its effect is proportional to its strength and it might happen that its effect is negligible, but there are no activation rules for the defeater's arguments.

Rule 3: Accrual of reasons

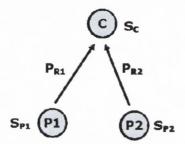


Figure 4.14 Two independent arguments supporting the same conclusion

Accrual of reasons occurs when a conclusion C can be inferred by two or more separate arguments, which makes the conclusion C stronger than the one based on only one of the reasons. Pollock denies the accrual of reasons for epistemological reasoning, in which an agent has to decide what to believe, while it might be true for practical reasoning, where an agent has to decide which action to take.

Pollock concludes that in epistemological reasoning the strength of a conclusion based on two or more separated arguments is equal to the maximum of each single supporting argument.

$$S_c = max(S_{p1}, S_{p2}, \dots, S_{pn}) \quad (4.4)$$

Trust is a reasoning that has the characteristics of both practical and epistemological reasoning. Trust is a practical reasoning, since it is driven by an action/decision to be performed. Therefore, in our semantic we do not reject the rule taken from Pollock, but we consider the rule limited only to a specific set of cases, not able to capture all the ways different arguments' strength is aggregated into one.

In our model we prefer to define a generic function of aggregation F_{agg} at each node, defined as follows:

$$F_{ag,g} \colon \mathbb{R}^N \to \mathbb{R}$$

The above formula (4.4) is a first example of function of aggregation, but we also include other possibilities. For instance, F_{agg} could be based on the use of an averaging approach, adequate when the different reasons have all to be considered in the final conclusion; or on the weakest principle, modeling a pessimistic way of aggregating the evidence; or on considering some factors optional and so forth. Examples of F_{agg} are provided in our implementation chapter 5. Of course each of this aggregation function has to be chosen in relation to the nature of the specific conclusion under analysis, and the choice of an aggregated function could involve a domain-specific analysis. The maximum-based approach is not adequate when the arguments on which the conclusion is taken are radically different in their structure. For instance, if the conclusion "Trust Mark" is based on two separated reasons, one being "Mark did well in the past" and "Mark has a high reputation", we prefer an average approach rather than selecting the maximum, since we do not want to lose information about one of the two argument, and we do not see any reason why an agent should use only the maximum reason instead of other approaches.

Rule 4: Role of defeaters

The problem is to estimate how the strength of an argument changes if supported or attacked by another. The only case to be considered is the undercutting defeater or supporter, the situation illustrated in the graph below:

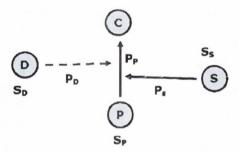


Figure 4.15 Defeaters and Supporters

The undercutting defeater D and the supporter S change the value of the plausibility P_p , and consequently the strength of the conclusion C.

Pollock treats defeaters with the theory of diminishers. Pollock's theory does not consider supporters that can increase the validity of an argument, that are explicitly inserted in our model. These kinds of supporters are not necessary for C to be asserted, but their presence makes the plausibility of the argument they support stronger.

In line with Pollock, C can be defeated or diminished by a value proportional to the strength of the defeater argument. The strength S of the supporting or defeating argument is important to determine how the plausibility P_p will be affected, and consequently C.

Regarding the role of supportive arguments, if they satisfy the rule of activation, the impact on the value P will be determined by the following function of support $P_p = F_s(S_s P_s, P_p)$ (4.5)

A defeater is always active and the impact over P_p is given by the function of defeating
$$\begin{split} P_p &= F_d \big(S_s P_s, P_p \big) \\ P_p &= F_d \big(S_s P_s, P_p \big) \end{split} \tag{4.6} \end{split}$$
A def

 F_d must satisfy the rules of diminishes introduced by Pollock and described in chapter 3 section 3.1.

4.6.1 Note on the computation

We now focus on the way argument status should be computed. The defeat status of the node has to be computed using the rules described in chapter II. Therefore, the computation is a recursion from the top nodes down to their premises, and then the premises of those premises and so forth.

The main difficulty in doing so is represented by the presence of circular paths, where, for instance, an argument attacks another by which its strengths depend. The critical paths are present in our trust-based reasoning. In our graph a first source of mutual dependency is represented by rebuttal defeaters, which are by definition bi-directional. At the very top of our graph there are two opposite arguments, Trust and Distrust, that are rebuttal defeaters.

The way to solve this dependency is the one suggested by Pollock. The resulting strength of the argument is derived by the comparison of the strength of the argument that are supporting the two conflicting arguments as described in chapter 3, end of section 3.1

A second common situation is represented by the graph showed below, analyzed again in section 3.1, chapter 3.

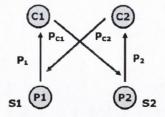


Figure 4.16 Circular Path

The results obtained in chapter 3 is that again the circularity is resolved by comparing the strength of the supporting arguments P_1 and P_2 .

What is important for our computation is that the computational procedures described allow us to apply a recursive computation that gives to the whole process a stage approach, in which the argument has to be computed from leaf nodes to roots nodes, represented by trust and distrust.

This reconciles the discussion performed in the previous part of the chapter with its formal representation.

4.6.1.1 Argumentation and Computation

The scope and meaning of our argumentation approach is to assess the plausibility value P, that tells us how strong the link between the output of such a computation and the strength of the premises is, i.e. at which degree the latter can be obtained by the former.

The plausibility study requires collecting facts that can undermine or strengthen the plausibility of the computation. The goal of the argumentation is to assess the plausibility of a computation, rather than performing the best computation among the possible ones. With best computation we mean the one that exploit the facts collected more correctly and with more precise output values. The sense of argumentation is that, whatever computation is used, there is a value of plausibility attached to it that has to be investigated, and that is essential to correctly estimate the strength of the premise.

If the computation used is simple or even naïf, its plausibility could be easily reduced, so that it becomes too weak to support trust scheme premises, but the same can happen for a more complex computation.

Certainly an agent may start from a computation and realize, after new pieces of evidence are available, that this computation is no longer valid, and it might decide to define a new computation to address the undercutting argument of the old one.

If the computation is more sophisticated and exploits a large set of facts and their mutual relationship, the value produced has a higher starting plausibility that may be reduced – or even increased- by external arguments but it results more solid and hard to be attacked.

Therefore, a more effective computation is obviously a better starting point, but even a simple computation, thanks to the plausibility study, has less possibility to produce incorrect results (true negative), that should be blocked by the plausibility study. The number of false negative should be reduced.

We underline how in general a more complex computation does not imply a more plausible value, and the opposite could be true.

The spirit of defeasibility is that any argument, from the simplest to the most complex, is a potential target for being defeated. This implies that, whatever computation is used, or whatever computation is re-formulated, the assumptions, facts and reasons behind it must be explicitly declared. It is the strength of all these factors at given epistemological states that defines at which extent a computation is valid. Complexity is not linked to plausibility by any positive or negative dependency. A simple computation might be more plausible than a complex one, if the extra complexity is based on facts and assumptions that can be easily defeated.

The defeasibility theory does not aim to show that a complex computation is more effective than a simpler one, but to suggest that plausible computations – simple or complex – should be considered more than implausible ones in a given epistemological state. If a computation is considered not plausible, results could be unpredictable.

The same computation, but in different contexts, may have different plausibility, resulting more effective in the context where plausibility is higher.

In this thesis we do not pursuit the best computation approach but we focus on the effectiveness of the plausibility study. Our goal is to show the beneficial impact on the computation of such study. Nevertheless, the plausibility study has the effect of suggesting a better computation, by addressing the effect of potential defeater/supporter in the computation directly.

This does not mean that we use naïve formulas to compute strength premises; the formulas we used have to be valid, in the sense that they show a predictive capacity regarding trust. Moreover, the formulas used in our model are well-known in computational trust models or grounded in social science studies accepted by literature.

The approach of extending the computation in order to address potential defeaters and supporters can lead to undesired results. The risk is to define a new complex formula that hides its assumptions or maybe whose assumptions are too complicated to be made explicit. This will affect the possibility that another agent – or the owner of the computation itself – attacks or supports the output of such a computation. A more complicated formula implies the risk of resulting "isolated" inside the mind of an agent rather than an argument cast in a dynamic discussion. The risk is to embed all the reasoning inside a formula. If this happens, the formula-based approach takes a direction that is opposite to the argumentation, a closed approach rather than an open and multi-agent one.

Second, a more complex formula has a plausibility to be studied. When an agent changes the computation attached to a trust scheme in order to address potential defeaters, it must make explicit the new facts and assumptions used, and define how to assess the plausibility of the new computation. Finally, the computation has to define how the plausibility changes if one of the inputs is missing.

The key point is that the elements of the computation and the way they are mixed **have to be understood and made explicit** in order to be part of an argumentation that can attack or support them.

4.7 Model's Contribution

In this section we briefly underline the main theoretical contributions and features of our methods.

Introduction of Defeasible reasoning and semantic formalization

The main theoretical contribution of our model consists in the introduction of defeasible reasoning in computational trust models. In particular, we formalize the notion of trust schemes – that make trust not only a (defeasible) reasoning but a reasoning initiated and conducted around recurrent patterns. We defined and formalized a semantic for defeasible trust reasoning extending and specializing Pollock's semantic.

Set of trust schemes

By saying that recurrent trust schemes are the backbone of our reasoning, we stated that trust is a recognizable expertise, with proper mechanisms we make explicit in the following chapter. Therefore, our model avoids risk of reductionisms or lack of meaning and it is able to justify its conclusions. The model does not only give a *form* to trust, as the majority of the model does, but it also provides a *content*.

Evidence Selection

Our model contains a strategy for evidence selection. An element of the application is selected as evidence if it can support the premise of a trust scheme. The selection of the pertinent

pieces of evidence is therefore linked to a generic procedure and it is not totally delegated to a process external to the trust model, usually represented by a domain-specific expert.

The role of the expert in our model is not to deliver a trust solution but only to help, if needed, the completion of a specific task requiring knowledge of a domain property, addressed by the critical questions inherent to each trust scheme. In other words, these elements can be seen as acting like trust instances, general trust schemes instantiated with domain elements. Instead of having a set of evidence that an expert recognized, trust instances replace the domain-specific expertise with a generic expertise of trust. On the other hand, trust instances keep a rich notion of trust encoded in the schemes. By doing this, we respect the requirement to not delegate the trust solution to an expert.

Conclusions

In this chapter we have introduced our method's design. The method is derived from the idea that trust is a form of defeasible reasoning. This reasoning is composed by recurrent arguments, generated by generic patterns called trust schemes. Each trust scheme has a defeasible nature: it might be attacked or supported in the light of new evidence.

Trust schemes are at the core of out methods: they initiate the reasoning by producing arguments pro or against a trustee, and by linking evidence and trust arguments they support an evidence selection strategy.

We described the stages of the method: the evidence selection, the computation and testing of the trust scheme and the final argumentation between the trust schemes' conclusions.

We provided a formalization of the concept of trust reasoning and trust scheme in terms of inference-graph, and we provided a semantic for computing the status of the arguments we implement in the following chapters. Our method, as any defeasible reasoning technique, has its natural application in a dialectic process between two parties.

Finally, we have underlined our contributions, represented by the introduction of defeasible reasoning and the definition of a list of trust schemes proposed in the next chapters. Trust scheme constitutes an explicit and defined trust expertise, which avoids reductionism approaches and lack of meaning in the method. Regarding the role of domain-specific expertise in the method, we showed how it plays an important but supportive role.

Chapter 5 Implementing the model

Introduction

In this chapter we describe the implementation of our model of trust based on defeasible reasoning. In order to accomplish this task, the following issues have to be analysed:

- 1 Define how each trust scheme generates an argument with a quantitative strength for a specific entity
- 2 Define the list of defeasible trust schemes, encompassing the way each trust scheme is mapped/computed over domain elements and a list of critical questions to assess its plausibility
- 3 Define a semantic for our trust-based defeasible reasoning, i.e. how to combine different trust schemes – and the corresponding arguments – into sustainable and logically consistent conclusions.

In the first section of the chapter we present the ranking-based computation used to quantify the strength of each argument generated by a trust scheme. The key idea is that every applicable trust scheme is computed for the entire population, and entities are ranked on the base of that value.

The strength of an argument for a specific entity is therefore proportional to the position in the ranking.

In the section we also briefly describe some basic computational tools needed to compute trust value and analyse ranking, mainly fundamental concepts of *ranking statistics*, the branch of non-parametric statistics that deals with *distribution free* set of data.

In the second section of the chapter we present our list of defeasible trust schemes. We greatly rely on our theoretical design of chapter 4 and on the state-of-the-art review of the previous chapter 3. The section provides the following:

- 1 The list of the trust scheme, the generic arguments to trust
- 2 The definition of their critical tests, i.e. rebuttals and undercutting arguments
- 3 Formulas and computational methods to convert a scheme into a quantitative value

The computational method presented is far from being the only possible and comprehensive, its main scope is to allow us to quantitatively evaluate the core hypothesis of this thesis.

In section 3 we present the mutual relationships among the schemes, used to combine arguments into a defeasible semantic.

Finally, in section 4 we provide the functions that compose our defeasible semantic. In particular, we provide:

- The function of conclusion f_c , that defines the strength of a defeasible conclusion
- The function of attack and support f_a , that defines how the strength of an argument changes in presence of defeaters or supporters
- The accrual of reasons
- The mutual relationships among our trust schemes, needed to aggregate the defeasible arguments into a final trust value

5.1 Ranking-based Computation

Our computational model is centred on the notion of rank. The method requires to compute different trust schemes for the set of entities whose trustworthiness needs to be studied. Each trust scheme generates a ranking on the set of entities. The strength of each trust scheme argument for a specific entity is proportional to its position in the ranking. Our convention is that the 1st position of the ranking represents the most positive one.

Rankings coming from different trust schemes are manipulated and combined with other rankings according to the status of the reasoning – supporter, defeaters, rebuttal, aggregation - generating new rankings among entities. Trust and distrust represent the two final rankings on which entities are judged. The situation is depicted in figure 5.1. The simple trust reasoning depicted is composed by three arguments, each of them with its ranking. Arguments 1 and 3 support directly trust, while argument 2 is an undercutter of argument 1. In order to compute the rank for argument T, we need first to modify the ranking of argument 1 taking into account the effect of argument 2.

For instance, let's assume that $R_1(x)$ is the ranking of the entity X for trust scheme 1, and $R_2(x)$ for trust scheme 2. A function of attack F_a is needed to understand how argument 2 affects argument 1. We remind how the function of attack defines how an argument changes when attacked by an undercutting defeater. Let's assume a simple linear function of attack, so that $R_1(x)$ is changed as follows:

$$R_1'(x) = R_1(x) + (R_{tot} - R_2(x))$$
 (5.1)

meaning that R_1 is increased (therefore the strength of argument is decreased) linearly by a quantity proportional to R_2 . Note that if R_2 is equal to 1 the effect of the defeater is maximum, while if R_2 is equal to R_{tot} the effect is null, as it is expected. Using the function of attack, R_1 is computed for each entity. The new ranking represents the trust scheme 1 after the application of the defeater argument 2.

Finally, a rank for T is computed by aggregating the new ranking of argument 1 with argument 3's ranking.

5.1.2 Positive and negative evidence

The strength of an argument is directly proportional to its position in the rank. The strength can be better quantified using statistical indicators as we describe in the next sections. Each ranking is mapped into a trust value in [-1,1] in the following way:

$$T_{v}(x) = \frac{\frac{R_{x} - \frac{R_{tot}}{2}}{R_{tot}}}{\frac{R_{tot}}{2}}$$
 (5.2)

Where R_x is the position of the entity in the ranking, assuming 1 as the most positive position. Positive and negative evidence are symmetric. If R_{tot} is the size of the rank, equal to the number of entities analysed, a positive argument is represented by a ranking, at least greater than $\frac{R_{tot}}{2}$. Positive evidence support a trust scheme S, while negative evidence support their opposite $\neg S$.

For instance, if S is the trust scheme activity, and entity X is ranked 100 and R_{tot} is equal to 1000, the entity has high activity (0.9) and low inactivity (-0.9). The sign of an evidence is an important information to activate mutual relationships among arguments.

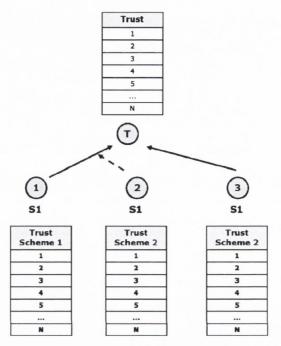


Figure 5.1 Rank-based strategy

We applied the ranking method in experiments I to IV described in chapter 6, choosing this ranking method for several reasons. Other strategies were possible in order to rank individuals value: we could have quantified the strength of a value for an individual i by looking at its relation with average and standard deviation of the distribution of values. We could have been to centre the distribution with a translation to zero and consider how many standard deviations a value differs from the average. This method allows quantifying how greater is a value a compared to b, while in the ranking strategy this information is lost: if the individual ranked 1st has a value that is 10 times greater than the 2nd or only marginally greater, the gap between the two ranking positions is always 1.

However, centring and dividing with the standard deviation has well-known drawbacks.

First, by normalizing the distribution, the presence of outliers can compress the values nullifying the information about the distance between values. Second, when different trust schemes have to be aggregated, no one can guarantee that the various distributions are comparable and aggregable.

No assumption can be done over the underlying distribution of values and a potential harmful mistake in our analysis could be to aggregate different sets of data that are not in the same form.

The rationale can be listed as follows:

- 1. By ranking all the individuals, we avoid the problem of outliers.
- 2. It gives all the different sets of data equal status, transforming actually each distribution into the same form, which was one of our priorities. On the contrary, by normalizing and centring each distribution, we would have introduced potential error in the way two values coming from different ranking are compared and aggregated. Another problem, local to each distribution, would be which semantic meaning to assign to a specific value on an absolute scale.
- 3. By using the fractional ranking, we have fairer individual rankings and more comparable to each other.

The alternative approach, based on centring and normalization, is applied in our earliest experiments in the context of Wikipedia reported in chapter 6. The computed value for each scheme is normalized in the interval [-1,1], centring the distribution to zero and by dividing by the standard deviation of the distribution. The main rationale behind this method is the fact that the information about the distance of two values is not lost.

5.2 Defeasible Trust Schemes

In this section we describe our set of defeasible trust schemes forming the content of our trust-based reasoning. We greatly rely on our theoretical design of chapter 4 and on the state-of-the-art review of chapter 3.

As described in chapter 4, a trust scheme is an *argumentation scheme* specific for trust. Therefore, a trust scheme is a generic defeasible reason to support trust or distrust. This section described the trust schema we identified, divided in seven overlapping macro-areas:

- 1. Time-based trust scheme
- 2. Trust-Scheme based on Information Sharing
- 3. Trust scheme linked to social role
- 4. Trust scheme based on activity analysis
- 5. Trust scheme based on outcomes
- 6. Trust scheme based on statistics and grouping

7. Trust scheme based on Game theory and Cognitive models and Risk

The source of our trust schemes is represented by the state-of-the art of computational trust models and social science literature. Our work is to extract, from the state-of-the-art, the underlying reasons that can represent a scheme. Some trust schemes haven been already proven to be effective- at least partially - in some trust computations. Some other trust schemes we identified represent a novelty: they are still grounded in social science but they do not have any computational model or evaluation regarding their validity.

Another novelty is represented by the investigation of the defeasible nature of each trust scheme, essential to support our argumentation.

Since trust schemes are presumptions, they are not definitive evidence for trusting an entity when they are positive, and vice versa when they are negative. Anyway, when a set of plausible trust schemes is properly aggregated, they can build a presumption stronger enough for the trustier to take a decision, as our evaluation shows. Trust schemes consider the same aspect from different angles, and their application helps practitioners to better design an effective trust computation.

Our list does not claim to be comprehensive, but it claims to represent a set solid enough to support a useful trust computation.

For easiness of reading, in the rest of the section we preferred to focus on the description of the assumptions of each trust scheme, the list of their critical questions and a description of how they can be computed without discussing our choices. A detailed discussion of our trust schemes, the rationale behind the set of critical questions chosen and the way each scheme is computed can be found in appendix C.

5.2.1 Time-based Trust Schemes

This class of schemes builds trust arguments using only information about time, usually temporal intervals between interactions.

It does not consider *what* was done during an interaction and - more important - *how* it was done. The focus is about *when* it happened.

These schemes represent a contribution to trust studies, since no evaluation of their effectiveness has been performed. They are present in some high-level computational models of trust and their importance is largely acknowledged both by social studies and by common sense. Time-based

trust schemes are: longevity, persistency, regularity and stability. Stability has been placed among time-based trust schemes since it focuses on how entities' properties change over time.

| | ble 5.1 Time-Based Trust Sch | |
|---|--|--|
| Presumption | Computation | Critical Questions |
| Longevity I trust an entity because of its longevity in the environment. Defeasible evidence of reliability and experience | $t_{last} - t_{present}$ is the time interval between the time t_0 – the <i>birth</i> of the entity – and t_{last} , the timestamp of the last useful moment | Is the choice of t_0 and t_{last} correct? How do you decide that an entity is part of the environment or not longer in the environment? Is longevity linked to a specific action or more actions/contexts? Is the environment competitive and selective? What is the cost of staying alive? Is there a positive or negative relationship between time and entities' abilities? |
| Persistency / Regularity Entities that persistently and regularly interact in the environment should be trusted, as a sign of accountability and reliability, while low persistency and regularity should be investigated as potential evidence for distrusting | An entity is persistent if, after dividing its lifetime in intervals of fixed time ΔT , it shows signs of activity in the majority of intervals. An entity is regular if the time interval between two consecutive interactions is relatively constant and not subject to high variance. See appendix C for details of the formulas used. | How is activity defined? Is the action frequent enough to collect enough data? Is the interaction supposed to be persistent and regular? Are there cycles? Is the time interval correct? |
| An entity having stable properties for a significant amount of time (typically until the present time) should be trusted. Stability gives you evidence that entity' status is well-established, mature and reliable. Instability can be seen as evidence that the entity still needs to evolve or correct/change its behaviour and abilities. | Given a set of properties $P=P_1$, P_2 ,, P_n , that are observable and quantifiable, we study the distribution of $D_p = P_n(t) - \mu(t)$. If we are more interested in a local increment we compute the function $C(t_1, t_2)$ that for a couple of timestamps t_1 and t_2 gives the difference between $[P(t_1)-P(t_2)] / P(t_1)$, and we compared the values with a threshold T | Is the entity active? Is the environment dynamic? Is ΔT significant? Is the entity young or in evolution? Is threshold T defining when an entity is not stable a reasonable choice? Are the properties P relevant? |

5.2.2 Trust Scheme based on Information Sharing and Social Role Information Sharing

This class of trust schemes bases its presumptions on information that members of the community share. As described in chapter 2, this is one of the two main computational mechanisms to compute trust, in the form of *recommendation* systems, *indirect experience* or *reputation*. All these concepts are build on information that is transferred between two members of the community (recommendations) or available to all the community (reputation).

| Table 5.2 Information-Sharing Trust Schemes | | | |
|--|----------------|--|--|
| Presumption | Computation | Critical Questions | |
| Indirect Experience | | | |
| An entity is trusted on the basis of what a third-party suggests | See Appendix A | (R _a is the entity giving a recommendation to T about B) Is the recommendation out of date? Is the recommendation pertinent to the context the trustee is interacting in? Is R _a trustworthy? Is R _a a good source of recommendation? Do T and B have compatible trust models? Is the recommendation first-hand? Do recommendations received by multiple sources agree? Is there a mechanism forcing entities to provide good recommendation rather than malicious? Is there a positive bias? Is there a link between agent R _a and B? Is the small world hypothesis tested? | |
| Reputation | 1: 1 | | |
| According to this trust scheme, entities are trusted for their | See Appendix A | Is the reputation system accepted? | |
| reputation value. | | Is there a positive bias? | |
| Reputation is a more compact | | Are name-changes easier and cost-free? | |
| value than a collection of recommendations; it is public and | | How is the reputation computed? | |
| therefore more transparent and | | Do the votes have different weight? | |
| accessible. | | Is multiple voting allowed? | |
| | | Is there a reciprocal vote effect? | |

Trust Schemes linked to Social Role

Trust Schemes linked to Social role base their presumptions on the connections that the trustee entity has in the environment. The schemes suggest that a trustee should be judged not in isolation but for the links and roles he has in the environment. Others may guarantee for him, or its public role/information may assure for him. The core pieces of information to be collected are: trustee's acquaintance, whom he is linked to and interacts with, his specific roles in the community, how easy it is to access and contact the entity, how transparent is the information he provided. As Carter wrote 'the reputation of an agent is based on the degree of fulfilment of roles ascribed to it by the society'.

| Tab | ole 5.3 Social-Role Trust Schen | mes |
|---|---|---|
| Presumption | Computation | Critical Questions |
| Authority (reference-based) | | |
| An entity is trusted according to the importance that other entities assign to him as revealed by indirect references | Analysis of the distribution of the references pointing to the trustee. PageRank approach | What is the meaning of the act of linking A by B in the domain? How frequent is the action of linking? Is it supposed to be repeatable? How big is the entities' space? How are entities connected? Small Worlds to be considered? Is the confounding variable age tested? |
| Connectivity | | |
| The presumption is that we trust an entity that shows high connectivity with many other entities in the community. Connections can be simple acquaintance, they do not mean an implicit judgement about entities as in the Authority trust scheme. Connectivity makes the entity more accessible and visible | Analogous to Authority. Signs of implicit acquaintance used to link users. | Does the evidence used really imply that A knows B? Is it possible to discern if their relations are positive or negative? |
| Popularity The trust scheme popularity states that an entity should be trusted because of its popularity | The scheme implies building a rank of a certain activity/service among the community | Is the market competitive? Confounding variable (price, time) |

| Visibility and Accessibility | | |
|--|--|--|
| An entity increases its trustworthiness by means of visibility and accessibility, clearly linked to the trust ingredient accountability | The quantification of accessibility/visibility usually requires testing the presence of information linked to the entity's identity and contacts, or when the entity was accessible. | Are the date verified? Are they verifiable? Is the measured time incorporating idle time? |
| Transitivity | | |
| The transitivity trust scheme presumes that trust can be transferred among trusted entities: if A trusts B and B trusts C therefore A trusts C | See Appendix A for examples in literature (Golbeck, Poblano, Sierra) | How long is the transitivity chain? Are there cycles? Is there evidence of referral trust ability of the elements composing the chain? Is the transitive chain valid? Is the small world hypothesis tested? Is the transitivity trust-based or social-based? Could we switch from acquaintance to trust? Do they relate to each other? |
| Information Provisioning | | |
| Members of the society should regularly contribute new knowledge about their friends to the society | The quantification of the useful information provided for the rest of the community (voting, contributions, feedbacks) | Same as for Visibility |

5.2.3 Trust Schemes linked to Activity

This group of trust schemes focuses on the activity of entities in the environment, i.e. what entities have done rather than when or how. It focuses mainly on quantitative aspects, not considering the outcome of an action. Information such as type of action performed, its context and topic are also considered, but not the outcome or the quality perceived by other entities. Rather, they are based on a classification of the type of actions coupled with an investigation of their complexity and pertinence to the context.

In the trust scheme pluralism, we focus also on the comparison of amount of activity of different users in the same context.

The basic presumptions is that an entity cannot be trusted if it is not active in the domain while the high activity could be regarded, under some conditions, as an evidence of a good health status, high reliability and success in interactions. These set of trust schemes are an alternative to the classical duo recommendation/past-outcomes.

| Table 5.4 Activity-based Trust Schemes | | | |
|--|---|--|--|
| Presumption | Computation | Critical Questions | |
| Pluralism | | | |
| Trust should be granted to what is the result of many entities cooperation/point of view, a collaborative activity of multiple entities and points of view while the opposite is an argument for distrusting it. Pluralism should guarantee an object to be less biased and more objective, representing direct reason to trust | Studying the basic statistical properties of the distribution of the contribution of each entity Y _n . See Appendix C for details. | Are the entities Y_n recognizable? How is the cardinality of the set Y_n ? What's the frequency F_a of the action A? Are the entities independent? Is the action A critical for X? Is the quantification of A plausible? Are contributions occasional or can they be repeated? Is it possible to cancel or revert the contributions? | |
| Activity Defeasibly, entities that are not active should be trusted only after further investigations. On the contrary, an active entity represents evidence whose plausibility should be analysed. | The basic computation implies to quantify the amount of activity of an entity in a domain. Usually, the computation ranks several indicators, each of them measuring some aspects of the activity. The indicators must be accessible, the action observable and somehow quantifiable. Uncertainty could affect this scheme more than others. The aggregation should be done by selecting the most adequate F _{agg} (see section 5.4 for a list of function of aggregation) | Is there more than one type of actions? Which is the relationship among the evidence found? Is there a hierarchy among the type of actions? Are there optional or compulsory actions? Is the environment competitive? Is the entity in a privileged position? Are entities actions in competition? How is the size of the market? How is the market trend? Which is the size of the entity? Does the entity need to be active to survive? | |

| Is activity supposed to be continuous? |
|--|
| (\Delta T=period of observation) |
| Is ΔT enough to quantify an entity |
| activity? |
| Which is the pertinence of the type of |
| actions? Is the activity pertinent to the |
| trust purposes of the |
| context/community? |
| Is the activity competent? |

5.2.4 Trust Schemes linked to outcomes

The most used computational trust mechanism, the past experience scheme presumes that an entity that acted well in the past, producing the expected outcomes, will continue to do so. Therefore, an entity is trusted for its past achievements as sign of its present ability to fulfil.

| Table 5.5 Outcomes-based Trust Schemes | | | |
|--|-------------|---|--|
| Presumption | Computation | Critical Questions | |
| Past-Outcomes | | | |
| An entity that acted well in the past, producing the expected outcomes, will continue to do so | | T_r is the trustier entity while T_e the trustee that in the past performed n times an action A with outcomes O_n . The action A have a frequency f_a . The period of observation is ΔT Are the outcomes objectively judged? Is the information out of date? Is the action A or the environment changed? Has trustee T_e changed? Is the number of collected outcomes enough? Was the outcome of action A affected by external constraints outside trustee controls? Does T_e have the motivation to accomplish the task? | |

Form a computational point of view, in chapter 2 we described various method for computing a trust value based on past experience that resembles a learning strategy enforced by feedback loop.

The scheme is regarded by some authors – such as the *Trustcomp* group [Tru04] - as an objective measurement of trust. Here we show its deeply defeasible nature that can sometime invalidate completely the scheme.

5.2.5 Trust Scheme based on prejudices and grouping

These set of trust schemes ground their assumptions on the statistical significance of some properties of the trustee compared to other entities or group of entities. The sociological motivation behind these trust schemes are the socio-psychological studies of Kahneman and Tversky [Tve74] while the use of *categorization* in Castelfranchi Falcone [Cas00] and the concept of *prejudice* in computational trust as described in section 4.1.4. Entities trust other entities on the basis of the categories they belong to (or they are supposed), or on the basis of similarities/dissimilarities with the trustier entity.

| Table 5.6 Prejudges- and Grouping-based Trust Schemes | | | |
|---|--|----------------------------------|--|
| Presumption | Computation | Critical Questions | |
| Similarity, Categorization an | d Standard Compliance | | |
| I trust what is similar to me Categorization I trust on the bases of the category an entity belongs to. Trust is assigned to category and passed on to its members Standard Compliance Entities showing properties significantly different from the average of their categories or from a defined standard should deserve further investigation and it is not prudent to grant them trust without further evidence. | A computation assessing the similarity between two objects, usually a n-dimension distance measures between a set of attributes describing the entity. The computation is similar to a Control charts-like computation. See Appendix C for more details. | Is the difference in positive or | |
| I trust entities similar to trustworthy entities | A notion of "trustworthy entity" is required | | |

The common idea behind these mechanisms is that trust can be transferred among similar entities/situations. Therefore, entities are pre-judged for belonging to a group. We remind how in digital world prejudice does not have a negative meaning but is *the mechanism of assigning*

properties to an individual base on signs that identify the individual as a member of a given group [Sab05].

These groups of trust schemes encompass *Similarity*, *Similarity to Trust*, *Categorization* and the Standard compliance trust scheme. They are all base on the same concept of similarity quantification, but the first analyses the similarity between the trustee and the trustier – therefore reflects a local point of view -, the second the similarity between the trustee and the *stereotype* of the trustworthy entity that the trustier build in its mind. The third scheme assess the similarity between the trustier and a group of entities and the fourth the similarity between the trustee and a standard –if any – that emerged or is defined in the entities community or in the mind of the trustier. The way of computing them and the critical tests associated are similar, so we describe the trust schemes all together.

5.2.6 Game theoretical/Cognitive Trust Schemes and Risk

These set of trust schemes consider opportunistic motivations that the trustier and the trustee may have in the situation, modelled as a game among rational players. The assumptions behind these trust schemes is that the trustee and the trustier are both rational entities that are trying to maximize their satisfaction and minimizing the effort spend.

Therefore, the understanding (or the presumption of knowing) the cost and benefit of the other entities produces an argument in favour or against trust.

These schemes should not be seen as a reduction of trust to a mere quantification of the utility that each party gains in the interaction, since this quantification includes cognitive reasons, such as the motivation of the entity.

Due to the kind of evaluation performed in the next chapter, these trust schemes were not investigated from a computational point of view in this work.

Anyway, we describe them here for giving a list of trust scheme more complete, providing the basic presumption and their critical questions and knowing how our description represents just a starting point.

The following schemes seem to be more useful in agent-based situation, more linked to outcomes and interactions than the scenarios used in our evaluation. Finally, sine risk is always incorporated in a trust-based decision; the Risk trust scheme considers the risk profile of the trustee as a reason to trust him/her.

| Table 5.7 | Game Theoretical-based T | rust Schemes |
|--|---|---|
| Presumption | Computation | Critical Questions |
| Common Goal/Situation/Risk | 444.44 | |
| If two entities share the same situation is more likely to help each others. Both the parties have interest in the situation and therefore they may merge their effort. | The scheme requires checking if the entities are in the same situation/share common goals | How can we be sure that the other entity is in our situation? Can he prove it? Does the trustee gain any advantage by the situation? |
| Cost/Benefit | | |
| a trustee entity should be trusted more if it has a very favourable benefit/cost ratio in the specific situation | A benefit/Cost analysis. Out of the scope of the thesis. See [Cah05] | The investigation of other motivations or reasons behind an interaction |
| Fulfilment | | |
| The trust scheme suggests trusting entities that are committed to fulfil the task assigned. The scheme does not mean that the entity will produce the desiderated outcome, but that he will do its best effort to try to achieve it. | Similar to past-outcomes. | |
| Risk profile | | |
| Entity A trust the trustee entity B if the risk profile of B is compatible with A's one. | Risk assessment. Out of the scope of the thesis. See [Cah05] | |

5.2.7 Trust scheme summary

This section provides a final summary of our list of trust schemes. The summary is depicted in table 5.8, where, for each set of schemes we report the defeasible presumption on which each of them is based

| Table 5.8 Summary of Trust Schemes | | | |
|------------------------------------|---------------------|--|--|
| Set of Schemes | Trust Scheme | Trustee's presumption | |
| Time-Based | Longevity | Trust entities with high longevity | |
| | Persistency | Trust entities acting persistently | |
| | Regularity | Trust entities acting regularly | |
| | Stability | Trust stable entities | |
| Information Sharing | Indirect Experience | Trust entities according to other's people recommendations | |
| | Reputation | Trust entities with high reputation | |
| Social-Role | Authority | Trust entities with high authority | |
| | Connectivity | Trust entities that are well-connected in the environment | |
| | Popularity | Trust popular entities | |

| | Visibility/Accessibility | Trust entities that are visible and easily accessible | |
|------------------|--------------------------|---|--|
| | Transitivity | Trust what your trusted entities trust | |
| | Information Provisioning | Trust entities that provide /share information | |
| Activity-Based | Pluralism | Trust entities or objects that are the results of many points of view | |
| | Activity | Trust active entities | |
| Outcomes-Based | Past-Outcomes | Trust entities that did well in the past | |
| Prejudge- and | Similarity | Trust entities similar to the trustee | |
| Grouping-based | Categorization | Trust an entity on the base of the category it belongs to | |
| | Standard Compliance | Trust an entity that satisfies a standard | |
| | Similarity to Trust | Trust what it is similar to what the trustee trusted | |
| Game-Theoretical | Common Goal | Trust an entity that shares similar goals, risks or situations | |
| | Common situation | | |
| | Common Risk | | |
| | Cost/Benefits | Trust an entity if it has a favourable benefit/cost ratio for the situation | |
| | Fulfilment | Trust entities that are committed to fulfil the task assigned | |
| | Risk Profile | Trust entities with a compatible risk profile | |

5.3 Mutual relationships among Trust Schemes

In the previous section we introduced our list of defeasible trust schemes that compose the trust-based reasoning and provide the arguments for the discussion. In order to complete the implementation of our defeasible reasoning, this section describes the following the mutual relationship among the trust schemes, partially contained in the critical questions. These mutual relationships will be used to reach conclusions that are logically consistent, by mean of the formulas described in the next sections, containing the implementation of our semantic.

5.3.1 Combining trust arguments: mutual relationships and argumentation

When a trust decision has to be taken, several pieces of evidence are usually gathered. In our model, the application of trust schemes produces a set of trust arguments, each of them expressed by a trust scheme applied to some specific elements of the domain. The final aggregation phase takes the set of available arguments and computes a final trust value.

One of the core issues of this work is how to perform this aggregation. Our hypothesis is that an aggregation should be performed by applying a defeasible semantic that takes advantage of

knowledge about the nature, strength and relationship of each arguments in relation with the others.

Usually arguments are mutually dependant and cannot be treated using a simple aggregation strategy. This is also linked to the defeasible nature of our specific reasoning: defeasible arguments supporting a common conclusion (trust or distrust) are likely to defeat or support each other. As describe in chapter 3, a simple aggregation is justifiable only when no information are available on the mutual relationships among evidence or it is known that no relationships exist. Our hypothesis is that an aggregation that is performed after an argumentation produces more efficient and consistent results.

Due to the nature of our method, the reasons why an entity should be trusted are stated in the trust schemes used. Since motivations behind a trust scheme are known, we can also explicit the mutual relationships among the schemes, by investigating if and how a trust scheme affects another. These mutual relationships – forming what in chapter 3 we called the *argumentation layer* - have been already partially defined in the critical questions paradigm.

Critical questions represent no more than arguments that attack or back a specific trust scheme.

Critical questions are tests totally defined by the trust scheme structure; they are inherent to a specific trust scheme. Critical questions usually generate a form of argumentation local to a trust scheme. Anyway, other critical questions define inter-schemes relationships. Some of them suggest considering other trust schemes value in order to check the plausibility of the one under analysis. In this section we make explicit those relationships already identified and we extend them introducing new ones. We remind the way different trust arguments may affect each other:

- 1. Rebuttal arguments
- 2. Support and Mutual support (undercutting)
- 3. Defeater and Mutual defeater (undercutting)

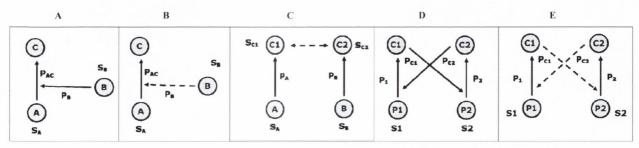


Figure 5.2 Mutual relationships among arguments: support (A), defeat (B), rebuttal (C), mutual support (D), and mutual defeat (E)

Rebuttal

As described in chapter 3 and 4, when two arguments are rebutting each other, the resulting strength of each argument is determined by the strength of the elements supporting each of them. Therefore, we do not need to define any attack function, since the strength of C_1 and C_2 (figure 5.2 c) is fully determined by S_a and S_b as follows:

- If $S_a > S_b$ then $C_2=0$ and $C_1=S_a-S_b$
- If $S_b > S_a$ then $C_1=0$ and $C_2=S_b-S_a$

Undercutting Supporters and Defeaters

In this situation, an attack/support function has to be defined. We remind how this is a function of the strength of the defeater/supporters S_b , the plausibility of the reasoning link attacking/supporting the argument P_b and the plausibility of the link between premises and conclusions P_{ac} (see figure 5.2 a or b). Therefore, if B attacks/supports A-C

$$P_{ac} = F_a(P_{ac}P_b, S_b) \tag{5.3}$$

Mutual defeaters and supporters

In the case displayed in figure 5.2d and 5.2e, arguments mutually support or defeat each other. The situation creates a circular path that, as described in chapter 3, is resolved by considering the equivalent graph obtained by removing the minimum set of links creating circular paths. This means that there are no needs to define new functions for this particular situation, that is reduced down into two separate situations.

5.3.2 Mutual relationships among trust schemes

In our rank-based method, each trust scheme produces a ranked piece of evidence. Our evidence are mapped into the interval [-1,1], and therefore they can be negative (lower half of the ranking), neutral or positive (upper part of the ranking). Our evidence strength is symmetrical: if S is the strength of a particular evidence, 1-S is the strength of an evidence in favour of the opposite assertion. A mutual relationship among two arguments may require the strength of an argument to have the proper sign. For instance, stability is supported by a positive (high) degree of activity, while is actually defeated by a negative (low) degree.

Table 5.8 shows the list of mutual relationships identified among the trust schemes, the sign displayed means that the evidence considered must have a positive or negative value for the rule to be applied. If sign is omitted the relationship is always applicable. This list does not claim to be comprehensive, but it is enough to study the impact of argumentation in comparison to a simple aggregation strategy. Moreover, the relationships identified are defeasible and they had a value of plausibility attached to them. In general, we could identify the following argumentation trends.

- The *time-based* trust scheme, such as *longevity*, usually has the effect of making the values of other trust schemes more plausible, for bad or good. Time-based trust schemes other than longevity add more detailed information.
- The *activity* trust scheme is crucial to many trust schemes: less or nothing can be said about an entity that is not active, while an active entity affects the conclusions of many others trust schemes.
- *Pertinence* is a key condition for activity and it may invalidate it, starting a cascade effect that can invalidate time-based and social-based trust scheme.
- *Social-based* trust schemes are affected by time, activity, but they have also some internal dependencies: reputation is affected by connectivity and transitivity.
- Motivation can invalidate the entire process, while a plausible value for outcomes-based trust schemes may confirm or contradict reputation values, activity, and longevity.
- Trust scheme such as *stability* may invalidate trust schemes that are time-based or that requires time to reach a solid value.

We note how these mutual relationships are not deductive but rather defeasible: a plausibility value is attached to one of them whose plausibility has to be studied.

Table 5.9 Mutual relationships among arguments

| T. Scheme A | Relationship | T. Scheme B | Plaus. | Comment |
|-------------|------------------|----------------|--------|---------------------------------|
| Activity+ | → (support) | Longevity+ | VH | A high longevity is |
| | | 53 | | strengthened by high activity |
| Activity+ | | Stability+ | Н | A high activity makes stability |
| | 1, 1 - 1 - 1 - 1 | | | a stronger evidence |
| Motivation+ | | Past-Outcomes+ | VH | A strong motivation makes |
| | | | | past outcomes more valuable |
| Activity+ | | Past-Outcomes+ | Н | Past-outcomes are a stronger |
| | | | | evidence if supported by a |
| | | | | high degree of activity |

| T. Scheme A | Relationship | T. Scheme B | Plaus. | Comment |
|---|------------------------|-----------------------------------|--------|--|
| Longevity+ | | Past-Outcomes+ | Н | A long history of good past outcomes is a stronger |
| Longevity+ | | Stability+ | M/H | Stability for a long period of time is a stronger evidence |
| Persistency+ | | Stability+ | Н | Stability is stronger if the entity is constantly active |
| Stability+ Longevity+ Persistency+ Activity+ | | Recommendation+ | Н | Recommendation is a stronger evidence if the entity did not change, if it is around since a long time and its activity high and constant |
| Activity- | - → (attack) | Longevity+ | VH | A lack of activity invalidates a high longevity |
| Activity- | | Stability+ | Н | A lack of activity weakens a high stability |
| Pertinence- | | Activity+/- | Н | A lack of pertinence weakens the degree of activity |
| Longevity- | | Persistency+ | Н | A lack of longevity makes a high persistency a less definitive argument |
| Persistency- | | Longevity+ | Н | A lack of persistency makes a high longevity weaker and empty |
| Longevity- | | Recommendation+ | Н | Recommendation are weak for young entity |
| Stability- | | Past-Outcomes+ | VH | A lack of stability invalidates past-outcomes |
| Motivation- | | Past-Outcomes+ Recommendation+ | VH | A lack of motivation makes reputations or a good history useless |
| Longevity- | | Past-Outcomes+ | M/H | Past outcomes are weaker if they refer to a short period of time |
| Activity- | | Past-Outcomes+ | | Past outcomes are weaker if the entity is not active |
| Past- Outcomes- | | Recommendation+ | | Recommendation is weaker if not supported by past outcomes |
| Stability- | | Recommendation+ | M/H | Lack of stability makes recommendation obsolete |
| Persistency- | | Stability+ | М | Lack of persistency affects stability |
| Past- Outcomes- | ← – → mutual attack | Recommendation- | | Low reputation and low outcomes are a stronger evidence against the trustee |
| Persistency+ | ↔ Mutual | Longevity+ | | High persistency and high longevity enforce each other |

| | support | | | |
|-----------|---------|-----------------|----|--------------------------------|
| Past- | | Recommendation+ | VH | If direct and indirect |
| Outcomes+ | | | | information agree, evidence is |
| | | | | stronger |

5.4 Semantic

In previous sections we provided the computational tools needed for implementing our system and a list of defeasible trust schemes used to generate arguments for our trust-based reasoning. In order to complete the implementation of the model, here we provide the functions required to compute the defeasible status of arguments. In particular, the following has to be provided:

- The function of conclusion f_c , that defines the strength of a defeasible conclusion
- The function of attack and support f_a , that defines how the strength of an argument changes in presence of defeaters or supporters
- The *accrual of reasons*, i.e. the function that defines how to compute the strength of an argument supported by n independent arguments of strength S_n .

5.4.1 Strength of the conclusion

As described in chapter III, a conclusion B is derived from premises A and a defeasible link with a generic formula:

$$S_b = F(P_{a-b}, S_a) \tag{5.4}$$

Where S_a is the strength of the premise A and Pa-b the plausibility of the reasoning link between a and b as shown in figure 5.3a.

How do we compute the strength of *C* in our rank-based implementation? The strength of a conclusion is computed according to the following two rules introduced in chapter 4.

- 1. $S_c = S_a P_{ac}$ iff argument a is a premises derived by facts, i.e. an initial argument of the reasoning
- 2. $S_c = \min(S_a, P_{ac})$ in all the other cases

The formula should have the following properties. The value of the plausibility P_{ac} should limit the value of S_c , while S_a should determine how close or far S_c is from the limit value set by Pac. In our implementation Pac is a real number in [0,1] where 1 represent the certainty and 0 represents the situation of complete implausibility. When Pac is equal to 1 the link is deductive

and the strength of C is equal to S_a : the strength of the conclusions if fully determined by the strength of the premises. When Pac is between 0 and 1, it should be decided how fast the plausibility reduce the strength of the conclusions.

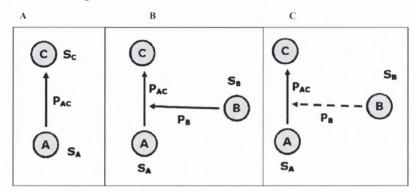


Figure 5.3 Function of conclusion (A), support (B) and attack (C)

An argument is strong only if both P_{ac} and S_a are strong, an argument is weak when plausibility is low or strength is low.

Both the rules a and b respect this criteria, but, as described in chapter III, rule 1 is based on a probabilistic model that would be erroneous to extend to the whole reasoning chain (conclusion becomes quickly unjustifiable), while rule 2 can be correctly extended to all the reasoning chain. The reason while the base-case follows rule 1 is because we avoid loosing information about the strength of the initial premises (i.e. the evidence).

In both the formulas, when P=1 $S_c=S_a$ when P=0 than S_c has the most negative value.

A factor n could be inserted in the rule 1 in this way:

$$S_n = S_n P_{nn}^n \tag{5.5}$$

The index n defines how fast a lack of plausibility will affect the strength of the conclusion. If n<1 the effect of a lack of plausibility is limited for high value of P and enlarged for low value of P and vice-versa. If, for instance, we set n=2/3, this means that a low value of plausibility has a bigger effect in proportion than an high value. In this way we have a more fine-grained ranking for the top entities and less-grained ranking for the low-ranked entities. This was because it is more important to discern which among two trustworthy entities could be the best choice instead of understand the best choice between two untrustworthy entities that in any case have both few changes to have trust granted.

5.4.2 Function of Support

As described above, our computational implementation is rank-based and the function of attack F_a will result rank-based. We remind the general shape of the function (A attacks B):

$$P_{a-c} = F_a(P_{a-c}, P_b, S_b)$$
 (5.6)

$$S_c = F(P_{\alpha-c}, S_{\alpha}) \tag{5.7}$$

More in details, the undercutting argument B will reduce the plausibility of the reasoning link between A and C, therefore modifying the resulting value of C. The strength of the argument A, S_a is not changed. How do we define F_a in our ranking-bases system? We divide the formula in two situations: supporting argument and attacking ones.

Supporting arguments

A supporting argument B should affect an argument A with the following desirable rules, anticipated in chapter III.

1. a supporting argument B support argument A only if it is stronger than A, i.e.

$$S_b P_b > S_a P_{ac} \tag{5.8}$$

Note how if S_b is negative and S_a is positive argument B cannot support A. The idea is that if argument A is stronger than B, A does not need any support from B and B cannot increase A If A is already stronger than B, it does not need any support and conclusion C remains fully determined by A. Finally, the argument A and B must have the proper sign for the specific supporting rule to be applied.

- 2. If rule 1 is satisfied, P_{ac} will be increased and therefore conclusion C will result stronger. The idea is that a supporter is an argument that help to make more plausible a reasoning-link. The magnitude of the increasing should be proportional to the strength of argument B (S_b) and the plausibility P_b . If Pb is high but Sb is low, the supporter has no effect since it is not based on solid premises, while if Pb is low, the supporters has no plausibility in increasing Pac.
- 3. The plausibility P_b of the supporting link will determine the maximum size of the increment $(P_{ac}$ can be increased proportionally to P_b P_{ac}) and the strength S_b determines how much of this increment will be assigned to Pa-c.

The formula chosen is the following:

$$P_{ac} = P_{ac} + c(P_b - P_{ac})S_b^n$$
 (5.9)

The resulting plausibility is increased by a value proportional to the difference between P_{ac} and P_b and the strength of S_b . The exponent n determines how fast the strength S_b will affect this increment.

We chose n=2 to emphasise the effect of very high value of S and reduce the (negative) effect of low value of Sb for the same reason described above, i.e. focusing on the more high-ranked entity. The constant c defines the maxim increment available. If S_b =1 (highest strength)

$$P_{ac} = P_{ac} + c(P_b - P_{ac}) \tag{5.10}$$

And c will determine the size of the maximum increment.

If c=1 then P_{ac} will be increased (at maximum) to P_b , while if $c=\frac{1}{2}$ P_{ac} can be increased (at maximum) to the average value between itself and P_b . We decide for an intermediate value c=3/4. In table 5.9 is represented the resulting strength of the conclusion S_c for various random values of S_a , P_{ac} , Sb and P_b .

Table 5.10 Function of Support analysis

| Case | Sa | Sb | Pac | Pb | Sc | Pac new | Sc new | Rank | New Rank | Rank |
|------|------|------|------|------|------|---------|--------|------|----------|------|
| 1 | 0.58 | 0.06 | 0.18 | 0.41 | 0.11 | 0.18 | 0.11 | 18 | 19 | 1.0 |
| 2 | 0.73 | 0.11 | 0.44 | 0.99 | 0.32 | 0.44 | 0.32 | 11 | 14 | 3.0 |
| 3 | 0.88 | 0.76 | 0.33 | 0.48 | 0.29 | 0.40 | 0.35 | 13 | 11 | 2.0 |
| 4 | 0.25 | 0.10 | 0.51 | 0.24 | 0.13 | 0.51 | 0.13 | 16 | 18 | 2.0 |
| 5 | 0.89 | 0.84 | 0.14 | 0.90 | 0.12 | 0.58 | 0.52 | 17 | 5 | 12.0 |
| 6 | 0.52 | 0.38 | 0.47 | 0.02 | 0.24 | 0.47 | 0.24 | 14 | 17 | 3.0 |
| 7 | 0.44 | 0.03 | 0.69 | 0.58 | 0.30 | 0.69 | 0.30 | 12 | 15 | 3.0 |
| 8 | 0.39 | 0.31 | 0.15 | 0.08 | 0.06 | 0.15 | 0.06 | 20 | 20 | 0.0 |
| 9 | 0.84 | 0.91 | 0.11 | 0.82 | 0.09 | 0.57 | 0.48 | 19 | 7 | 12.0 |
| 10 | 0.64 | 0.35 | 0.32 | 0.87 | 0.21 | 0.41 | 0.26 | 15 | 16 | 1.0 |
| 11 | 0.99 | 0.12 | 0.37 | 0.27 | 0.37 | 0.37 | 0.37 | 8 | 10 | 2.0 |
| 12 | 0.93 | 0.16 | 0.51 | 0.47 | 0.48 | 0.51 | 0.48 | 6 | 8 | 2.0 |
| 13 | 0.38 | 0.49 | 0.99 | 0.25 | 0.38 | 0.99 | 0.38 | 7 | 9 | 2.0 |
| 14 | 0.68 | 0.92 | 0.90 | 0.97 | 0.61 | 0.95 | 0.65 | 4 | 3 | 1.0 |
| 15 | 0.87 | 0.13 | 0.80 | 0.14 | 0.70 | 0.80 | 0.70 | 2 | 2 | 0.0 |
| 16 | 0.41 | 0.75 | 0.82 | 0.14 | 0.33 | 0.82 | 0.33 | 9 | 12 | 3.0 |
| 17 | 0.81 | 0.27 | 0.93 | 0.51 | 0.75 | 0.93 | 0.75 | 1 | 1 | 0.0 |
| 18 | 0.71 | 0.27 | 0.90 | 0.83 | 0.64 | 0.90 | 0.64 | 3 | 4 | 1.0 |
| 19 | 0.86 | 0.73 | 0.59 | 0.24 | 0.50 | 0.59 | 0.50 | 5 | 6 | 1.0 |
| 20 | 0.64 | 0.07 | 0.51 | 0.10 | 0.32 | 0.51 | 0.32 | 10 | 13 | 3.0 |

Average Difference:

The last three columns of table 5.10 are: the rank without supporters (column Rank), the rank after the application of the supporting argument B (column New Rank) and the last column displays the absolute difference among the two rankings. The average difference among the 20 cases is 2.8, that is statistically significant as a paired t-test can show. This means that the effect of supporter can be decisive in the final ranking.

The impact of supporter affect mainly to intermediate cases, that benefit from the application of supporters. If a case is already very strong, the role of the supporter is limited. Cases in which the role of the supporter is bigger (bigger gap between Pac and Pb and high S_b) increase their ranking (such as case 9 or case 5).

Referring to table 5.10, the two strongest cases (17 and 15) are the strongest in both cases, and the differences between the two rankings among the top entities are lower than the other cases.

For instance, case 5 gains 12 positions from 17° to 5° due to the effect of an high supporter with P_b equal to 0.9. The same stand for case 9 from 19° to 7° position.

5.4.3 Function of attack: defeating arguments.

Argument B may attack the link between premise A and conclusion C, by reducing the plausibility of the link, in some cases to a null value, meaning that the argument is completely defeated and disappear from the current state of the reasoning.

As described in chapter III, argument B can completely defeat conclusion C only if P_b is greater than P_{ac} . In the generic case, the strength of the conclusion will be reduced. There is no case in which the presence of the defeater has no effect. This introduces an asymmetry: a supporter can be neglected if too weak but a weak defeater cannot. Supporters must be stronger than the supported argument to be effective – remember Pollock's observation about the accrual of reasons: a weaker argument does not add anything in support to a stronger one -, while defeater are always effective since they contradict statement that – stronger or weaker – goes in an opposite directions.

With an analogous reasoning we define the following formula:

$$P_{ac} = P_{ac} - \frac{1}{4} P_b (S_b)^{\frac{5}{2}}$$
 (5.11)

According to this formula, P_{ac} is reduced of a quantity proportional to the plausibility of the defeater P_b and its strength S_b . We chose the constant c equal to $\frac{1}{4}$. Note that when $S_b=I$ (strongest defeater), $P_{ac} = P_{ac} - \frac{1}{4}P_b$

If P_{ac} has a negative resulting value, the argument A is totally defeated. Table 5.10 shows an experiment analogous to the one described for the supporting argument. Now the situation is depicted in figure 5.3c. The value of conclusion C with or without the presence of the defeaters has been ranked. The difference between the two rankings is displayed in the last column.

Defeaters have a bigger impact on the ranking, since some arguments are totally defeated when P_{ac} is reduced to zero. In table 5.11, 5 cases are totally defeated, and the average differences of the two rankings is 3.9 against a supporters' difference of around 2.

In particular, very high cases lost many positions after the application of defeaters. If the highest case (4) remains the highest in both the situation, the second and third case (7 and 6) lost 6 and 9 positions. For instance, case 7 is attached by a defeaters with Pb=0.91 and Sb=0.7, that reduces the plausibility P_{ac} down to 0.06.

On the contrary, argument with weaker defeater takes advantage. For instance, case 12 gains 6 positions after the application of a weaker defeater than the others, that slightly reduces its Pac from 0.81 to 0.76.5If we want to emphasize the effect of defeaters, making them faster to act, the coefficient C (now $\frac{1}{4}$) could be increased. C regulates the impact of the plausibility value P_b of the defeater. We choose $\frac{1}{4}$ to have a dual situation with previous selection. The exponent regulates the effect of the strength of the defeater S_b over the conclusions. If n is increased, higher vale of S_b will have a stronger effect.

Table 5.10 and 5.11 shows how the formula introduced have a statistical significance, the average difference of 2.8 (supporters) and 3.9 (defeaters) are long-term average obtained over 1 000 samples.

Table 5.11 Function of Attack Analysis

| Case | Sa | Sb | Pac | Pb | Sc | Pac new | Sc new | Rank | Rank Def | Rank |
|------|------|------|------|------|------|---------|--------|------|----------|------|
| 1 | 0.85 | 0.83 | 0.18 | 0.67 | 0.15 | 0.00 | 0.00 | 15 | 19 | 4.0 |
| 2 | 0.90 | 0.67 | 0.19 | 0.68 | 0.17 | 0.00 | 0.00 | 13 | 19 | 6.0 |
| 3 | 0.15 | 0.79 | 0.42 | 0.58 | 0.06 | 0.00 | 0.00 | 17 | 16 | 1.0 |
| 4 | 0.78 | 0.47 | 0.99 | 0.08 | 0.77 | 0.97 | 0.75 | 1 | 1 | 0.0 |
| 5 | 0.76 | 0.33 | 0.38 | 0.33 | 0.29 | 0.32 | 0.24 | 6 | 4 | 2.0 |
| 6 | 0.62 | 0.66 | 0.53 | 0.63 | 0.33 | 0.19 | 0.12 | 3 | 9 | 6.0 |
| 7 | 0.96 | 0.70 | 0.59 | 0.91 | 0.57 | 0.06 | 0.06 | 2 | 11 | 9.0 |
| 8 | 0.32 | 0.25 | 0.66 | 0.02 | 0.21 | 0.66 | 0.21 | 8 | 5 | 3.0 |
| 9 | 0.55 | 0.11 | 0.37 | 0.07 | 0.21 | 0.37 | 0.21 | 10 | 6 | 4.0 |
| 10 | 0.73 | 0.47 | 0.25 | 0.49 | 0.18 | 0.09 | 0.07 | 11 | 10 | 1.0 |

| 11 | 0.94 | 0.57 | 0.32 | 0.66 | 0.31 | 0.04 | 0.04 | 4 | 13 | 9.0 |
|----|------|------|------|------|------|------|------|----|----|-----|
| 12 | 0.57 | 0.94 | 0.30 | 0.86 | 0.17 | 0.00 | 0.00 | 12 | 19 | 7.0 |
| 13 | 0.45 | 0.07 | 0.61 | 0.61 | 0.27 | 0.60 | 0.27 | 7 | 3 | 4.0 |
| 14 | 0.41 | 0.75 | 0.41 | 0.02 | 0.17 | 0.39 | 0.16 | 14 | 7 | 7.0 |
| 15 | 0.42 | 0.58 | 0.23 | 0.48 | 0.10 | 0.02 | 0.01 | 16 | 15 | 1.0 |
| 16 | 0.67 | 0.31 | 0.04 | 0.32 | 0.03 | 0.00 | 0.00 | 19 | 19 | 0.0 |
| 17 | 0.44 | 0.06 | 0.13 | 0.80 | 0.06 | 0.12 | 0.05 | 18 | 12 | 6.0 |
| 18 | 0.73 | 0.55 | 0.29 | 0.25 | 0.21 | 0.19 | 0.14 | 9 | 8 | 1.0 |
| 19 | 0.63 | 0.33 | 0.48 | 0.03 | 0.30 | 0.48 | 0.30 | 5 | 2 | 3.0 |
| 20 | 0.12 | 0.06 | 0.09 | 0.28 | 0.01 | 0.09 | 0.01 | 20 | 14 | 6.0 |

Average Difference:

3.9

5.4.4 The accrual of reasons: the Aggregation of Different Rankings

As described in the chapter III, each argument has a function of aggregation F_{agg} , used to aggregate different distinct reasons to support an argument. Each reason has a different independent nature from another one, so that they cannot be compared straightforward. Pollock in his semantic define a single function equal to the max of the strength of the single reason. In our ranking based methods the strength of a piece of evidence is its position in the ranking for that specific indicator. The basic aggregation strategy is sum all the rankings into one, the only possible strategy if nothing is known about the different elements to be summed. In absence of any known meaning or relationship among the ranking, there is no reason to perform a strategy different from an aggregation, since it would introduce an arbitrary bias. When more information is known, such as (i) the plausibility of each ranking, (ii) the relative importance or (iii) the interrelationships among them, different F_{agg} could be more effective. We note how we are dealing with the function F_{agg} belonging to each node of the reasoning graph, not with the function of attack or conclusion proper of the argumentation process. Here the problem is how distinct pieces of evidence - whose mutual relationships are already taken into account and the defeasible argumentation, if any, has been already performed - related to the same argument can be combined in order to quantify the strength of the argument.

Given two ranking R_1 and R_2 , both ordered so that 1 is the most positive score, we define simple constructs that we will use in our evaluation and that can be generalized for n rankings:

Aggregated ranking

$$F_{agg} = R_{tot} = R_1 + R_2 \tag{5.12}$$

This combination is to be used in absence of any extra information about rankings plausibly, relative strength or interconnections. Note how the method could be the most adequate one if we know that each piece of evidence has the same strength or if it is known that they are completely independent, or in situation of full ignorance where nothing is known about they relative importance.

Linear combination ranking

$$F_{agg} = R_{tot} = aR_1 + bR_2 \tag{5.13}$$

This combination is to be used when the relative strength of the ranking is known. It is usually hard to estimate the size of *a* and *b*. A wrong choice of the parameters could affect the computation by giving to a ranking too much importance. Anyway, the method remains the most simple way to give to a ranking more importance than another one. If it is known that a ranking represents evidence that in the context is stronger and more plausible than another, the method could be used. Nevertheless, knowledge collected in the domain maybe clearly indicates the use of that ranking.

OR operator

$$F_{a,q,q} = R_{tot} = \min(R_1, R_2)$$
 (5.14)

The OR combination take as the final ranking the most positive one. This combination is used when it is known that not all the two evidence are needed, but only one suffices. A typical situation is the one in which an entity has many ways of accomplishing a task and all leads to the same results. The aggregation models also an optimistic vision of the final value.

$$F_{agg} = R_{tot} = \max(R_1, R_2)$$
 (5.15)

The AND combinations takes as the final value the most negative one. This means that in order to have a high score, an entity should have all the scores high. This strategy is used to model situation in which it is known that all the evidence are needed, no one can be neglected and the final value is strongly affected by only one negative value. As an example, you may consider the vital organs of the human body or the strength of a chain that are all determined by a single

component. The strategy models also an extreme pessimistic disposition, in which a trustee is judged on the base of its weaken point.

Compulsory and optional rankings

if
$$R_1 < R_2$$
 then $F_{aaa} = R_2$ else $F_{aaa} = (R_1 + R_2)/2$ (5.16)

This combination has the following meaning: R_2 represents a ranking that is compulsory in the final aggregation, while R_2 is an optional one. The total rank is solely R_1 when R_2 is worse than R_1 , while is the average of the two when R_2 is better than R_1 . In other words, R_2 can increase the value of the final ranking, but not decrease it. The combination models a situation where R_1 is a requested piece of evidence while R_2 represent a plus, that is not needed in order to have a high score but can help. When R_2 can only decrease the score, the computation is performed in the opposite direction.

Ad-Hoc aggregation

When enough information is available, F_{agg} can be defined as an ad-hoc computation. In our experiments we do not use any ad-hoc computation for F_{agg} . In any case, we note how any assumption underlying the computation must be explicitly stated in order to be defeated or supported by other arguments.

5.4.5 Final Aggregation

As described in chapter 4, the result of the argumentation is a new set of arguments, whose new status has been determined by relationships with other schemes.

This new set of value contains a set of positive evidence P, a set of negative evidence N, a set of neutral evidence A representing the surviving arguments after the application of the defeasible semantic.

Usually, some arguments have been defeated and they disappear from the final computation of the trust value at the current epistemological state.

The final trust value is now computed with a simple aggregation. Note how we aggregated arguments surviving the defeasible reasoning – justifiable arguments – with a computed strength. Since at the top of the reasoning there are two rebuttal arguments -trust and distrust- the final

trust value will be the difference between the sum of all the arguments supporting trust subtracted by the sum of all the arguments supporting distrust. Therefore the final value is simply:

$$T_v = \sum T_s^+ - \sum T_s^- \qquad (5.16)$$

where trust results a positive number and distrust as a negative one.

Conclusions

In this chapter we introduced the implementation of our trust system based on defeasible reasoning. In section 1 we presented our rank-based method for computing our trust reasoning. In this method, the strength of each argument for an entity is proportional to its position on the ranking. The support and attack relationships are functions that manipulate rankings and produce other ranking. We introduce the required basic statistical tools to study ranking distributions and we provided the rationales behind this choice, being mainly the distribution-free nature of the method and the limited roles of outliers' entities.

In the second section we presented the set of trust schemes used to initiate our trust-based reasoning. Our sets of trust schemes have been derived from multidisciplinary trust study and some of them have been used in previous trust computation. The list does not claim to be comprehensive, but it claims to be a set solid enough to sustain a meaningful and useful trust computation.

For each of them we have discussed their defeasible nature, providing a set of critical questions to be considered to assess its plausibility, coming from challenging the assumption on which each scheme is based. Again, critical questions are not meant to cover all the possible arguments defeating or supporting the specific scheme, but rather a set solid enough to make the usage of the scheme more efficient. We also provided, for each trust scheme, a computation able to generate a base-value on which the reasoning can start. Our aim was not to provide the most efficient computation, but rather focusing on the assumptions and required inputs each computation needs.

We have also provided a set of mutual relationships among the trust scheme. This set of relationships do not claim to be comprehensive but it provides a solid base to evaluate the impact of an argumentation strategy on the issue of aggregating different pieces of evidence linked with mutual relationships.

In the conclusive part of this chapter we provided the implementation of the functions required to compute the defeasibility status of our trust arguments. Using our rank-based approach, we have defined the strength of a conclusion (function of conclusion), and the function associated with supporting arguments and undercutting defeaters (function of attack and support), and we provided various implementations of the aggregation function F_{agg} , to be used when different reasons have to be aggregated into a final decision.

Chapter 6 Evaluation

Introduction

In this chapter we describe the evaluation of our computational model of trust. Evaluating a trust model means to quantify the ability of the model to give useful predictions about entity trustworthiness. A prediction has the form of a trust value attached to an entity.

In order to understand if our trust values are accurate, a second set of data has to be collected for comparison. In order to be valid, the second set of data should exhibit some properties we described in our introductive chapter. For the sake of completeness, figure 6.1 depicts again the comparative evaluation strategy, which will be used in all of our experiments.

The evaluation of the effectiveness of our trust values is therefore a comparative analysis between two sets of data, generated independently, where the second one is the benchmark. The more our computed set of values is similar to the second one, the more our predictions are effective. Following this criteria the following aspects of the model will be evaluated: the effectiveness of each trust scheme, the impact of the critical questions (i.e. defeasibility), the impact of an argumentation-based strategy against a simple aggregation one and the overall accuracy of the results.

The chapter describes seven different experiments conducted in the context of an online web-community and on the Wikipedia project. The chapter is organised as follows: the first section describes the hypotheses to be verified, the second section describes the experiments performed while the third section contains the conclusive discussion of this thesis.

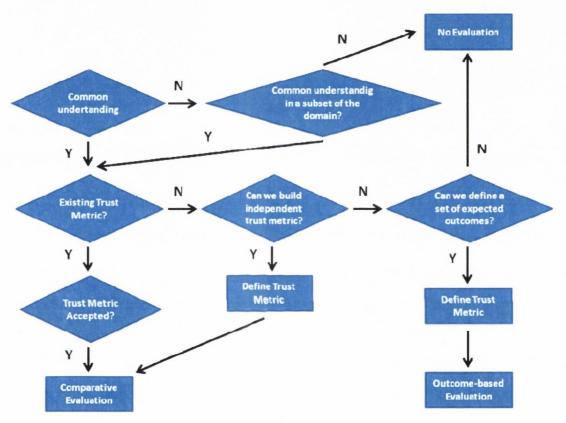


Figure 6.1 Evaluation strategy

6.1 Aspects under evaluation

As described in our introductive chapter, our evaluation strategy should focus on the following three aspects:

1. (Hypothesis 1) The efficacy of the critical question, i.e. the impact of defeasibility over the computation of trust scheme. The hypothesis to be verified is the following: after a critical question analysis, the trust scheme considered plausible should be effective while defeated trust schemes should not be effective. Bad arguments are therefore cut – reducing the number of true negative – and good arguments are strengthened – increasing the number of true positive. When a trust scheme computation results not plausible, there

are two possible cases: the trust scheme is discarded or it could be re-computed in a more plausible way. In the first case we expect that the trust scheme, if applied, would produce weak trust value –and therefore it was good to discard it – while in the second case we expect better results than the original computation and effective on an absolute scale.

- 2. (Hypothesis 2) The aggregation vs. argumentation. The hypothesis to be evaluated is the following: an argumentation-based aggregation that considers the mutual relationships among different evidence, their logical consistency and their relative strength leads to more accurate results than a simple averaging-like aggregation strategy.
- 3. (Hypothesis 3) The hypothesis to be evaluated is the following: the overall results of the method are of high quality. Note how the first two hypotheses are not contained in this one, since they focus on the benefits that defeasibility brings to trust computation, while this hypothesis focuses on the quality of the results on an absolute scale. The hypothesis is verified if results obtained show a high accuracy of predictions, meaning that our set of trust schemes and our defeasible argumentation are an effective tool for mining trust.

6.1.1 List of Experiments

We present seven experiments conducted in the context on an online web community of traders (FinanzaOnline.it) and in the context of the Wikipedia project. Here we provide the list of experiments and the hypotheses they evaluate. A detailed description of each dataset is provided in the rest of the chapter.

Table 6.1 List of Experiments

| Experiment | Context | Туре | Goal: (Trustworthiness of) | Hypotheses under evaluation |
|------------|---|-----------------------------------|----------------------------|-----------------------------|
| 1 | FinanzaOnline.it (advanced evidence) | Full application of the method | Forum members | 1,2,3 |
| 2 | FinanzaOnline.it (subset of evidence) | Full application of the method | Forum members | 1,2,3 |
| 3 | FinanzaOnline.it | Past-Outcomes Trust Scheme | Forum members | 1 |
| 4 | FinanzaOnline.it | Info-Provisioning Trust scheme | Forum members | 1 |
| 5 | Wikipedia | Full application of the method | Wikipedia Articles | 1,2,3 |
| 6 | Wikipedia | Full application of the method | Wikipedia Authors | 1,2,3 |
| 7 | Wikipedia | McGuinness metric | Wikipedia Articles | 1 |

6.2 Analysis of the Experiments

6.2.1 FinanzaOnline.it, trust-based mining of a large on-line Community

FinanzaOnline.it is the biggest Italian online portal about finance and trading. The experiment proposed in this section analyses the activity of this popular forum in order to reason about the trustworthiness of its members. The experiment challenges the well-known problem of virtual identities' trustworthiness, in a digital world where changing a virtual identity is effortless and cost-less, compromising trust and reputation.

finanza®nline

The *FinanzaOnline.it* [Fin08] forum is the oldest and the largest Italian forum dedicated to trading, in terms of both daily visits and number of members. It was opened in 1998, and it counts about 84 000 (up to June 2008) registered users, of which about 10 000 are daily habitual users, whit an average of 4 500 users connected, about 7 million archived messages in more than 200 000 threads. The forum is dedicated to stock market trading, mainly Italian but also American and other European markets. It contains three main sections: trading, news and finance in general and a section for free discussion and topics other than finance.

The discussion of this experiment is organized as follow: in the next section we perform the preliminary investigation of the domain, then we present the stages of our method: the matching of trust schemes over an application model (evidence selection), the critical questions analysis and the final argumentation.

Investigation of an existing trust-ranking

As stated above, the aim of the experiments conducted in the context of FinanzaOnlie.it is to assess the trustworthiness of virtual members. Scope of the preliminary investigation is to define a set of independent data to be used as a comparison with our results. Our starting point is represented by the following two questions: is there an accepted notion of trust? Is it contained in some pre-existing metric?

FinanzaOnline.it forum is a place where small traders or professionals share comments, suggestions, analyses of companies on the stock market. The aim of the community is clear: maximization of the profits by sharing the most useful and accurate information.

The most trustworthy members are the ones that provide useful information for trading, in the form of explicit suggestions, analysis, comments or news.

Given this, we could define an outcome-based definition of trust by stating that trust can be measured on the number of successful predictions/suggestions given by each member.

This definition of trust could be used, but it is actually time-consuming and, without a manual checking, it is hard to bind suggestions to outcomes. We performed this analysis on a limited scale in experiment III, here we need to follow a more scalable and feasible strategy.

A recommendation system is already in place inside the forum. Could it be used as a comparative set?

The answer is negative. By analysing a series of messages posted in the faq and suggestions section of the forum, it was clear how the existing recommendation system is not regarded as useful or fair [Fin08]. The system was regarded as "a joke" by many members of the community, suffering from the problem of mutual exchange of positive feedback among friends, with very few negative feedbacks mainly based on personal issues rather than trading issues. We performed a more quantitative evaluation of the system, described in experiment II, discovering how the system does not carry useful information nor does it help to discriminate good or bad users.

We decided to set up a pool directly inside the forum in order to investigate the presence of a stable and recognized reputation ranking among the members. The question proposed to the community was:

"Which member of the community do you consider the more trustworthy and reliable, considering the contribution he gave to you/the forum in form of accurate suggestions, information or analysis?"

The poll was opened for 50 days, receiving almost 1 500 answers. The poll was anonymous — without the possibility of double voting — and the users could express up to three preferences. Many users posted additional signed comments and motivations for their vote. The poll was actually a series of polls that reduced the number of candidates at each turn, until we reached a complete ranking. The results of the poll showed a very clear consensus over the most trustworthy entities. According to the poll results, we divided the members in tiers. All the members are also ranked inside each tier.

The first tier contains the most 30 trustworthiness entities, covering the 88.5% of the preferences, a second tier contains 70 members and the third tier the rest of the members for which we received a rank. We collected votes for about 180 members. We did not collect any vote for the other members, considered therefore outside the rank. The first tier is divided in turn into a subtier composed by the first 10 entities, representing almost three quarters of the preferences. The vote distribution clearly shows the strength of the community consensus and supports the use of this rank as a valid independent comparison. Our experiments are focused on the tier 1 entities.

Table 6.2 – pool results

| Number of Members | Cumulative % of preferences (% of votes in each tier) | |
|-------------------|---|---------|
| 1-10 | 74.9 | Tier 1a |
| 1-30 | 88.5 (13.6) | Tier 1 |
| 31-100 | 97.8 (9.3) | Tier 2 |
| 100-176 | 100 (2.2) | Tier 3 |
| Others | 0 | Tier 4 |

Application model

FinanzaOnline.it is developed using the vBullettin software, currently the version 3.6.8. Since it represents the most used software for developing online forum, it gives to the following application model a degree of generality. In fig. 8.3b the UML model of the application is depicted centred on the class Member. A member activity is mainly the provisioning of messages that are grouped into threads, linked to a topic, usually a company on the stock market. Threads are grouped in macro-areas: blue chips (biggest companies on the stock market), small caps, a section for free chatting and a section for FAQ sent to the forum administrators.

Each member has a public profile where they can publish personal information, including some extra contacts like email, skype or msn, the reputation score, statistics about their activity in the forum. Members can also vote each other using the internal recommendation system.

Each message can contain an attachment, and it can quote other messages. Each member can define a list of friends or a black list. Messages of members in the personal black list are ignored and not displayed to the user who defined it. Members can also send private messages and receive them into a private mailbox, not accessible and therefore not used in our computation.

A member can also open a pool, by defining a question and the set of answers; have a public virtual portfolio linked to an internal game at which each member can compete. A general rank and all the transactions performed by each member playing the game are available publicly.

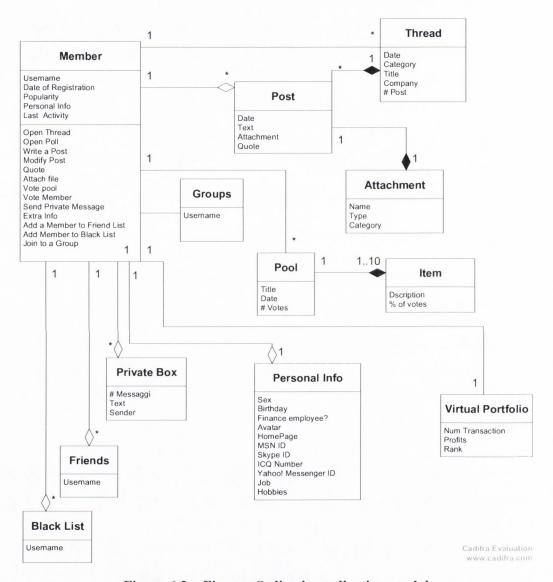


Figure 6.2 – FinanzaOnline.it application model

FinanzaOnline.it maintains a public archive of previous messages and threads. The archived messages cover around 80% of the whole activity of the forum, starting from 1998.

Using the archive we collected around 6 876 564 messages in more than 180 000 threads to be used for our computation.

6.2.2 Experiment I - Full application of the Method

Evidence Selection phase

We start our experiment by matching our trust scheme on the data available on the application model of figure 6.2. We report our results in table 6.3, where for each application data we analyse which trust scheme can be sustained.

Table 6.3 Matching data and trust scheme

| Data | Sign of | Trust Scheme | Data reliability | Comment on the TS |
|------------------|-----------------|--|------------------|------------------------------|
| Profile Info | | | | |
| Date of | Presence in | Time-based scheme such | Certain | |
| Registration | the system | as longevity | | |
| Popularity | NOT USED | NOT USED | Certain | The internal |
| | | 27 | | recommendation system is |
| | | | | not considered reliable |
| Personal Info – | Visibility, | Visibility/ Accessibility/ | Medium/Low | If contacts details – like |
| Contacts | accessibility | Social Network | Info are hard to | skype, msn – are valid, they |
| | | | be verified | are evidence that member |
| | | | | wants to be more visible |
| | | | | and accessible |
| Last Activity | Time | Longevity | Certain | |
| Friend List | Social | Visibility, Connectivity, | High, but rarely | |
| | Connection | Transitivity | specified | |
| Black List | Social | Visibility, Connectivity, | High, but rarely | |
| | Connection | Transitivity | specified | |
| Banned Status | Competence | Competence | Low – highly | Users are banned for bad |
| | | | controversial | behaviour |
| Thread | | | | |
| Open Thread | Activity | Activity | Certain | Open a discussion vs. |
| | | 3 TO 10 TO 1 | | posting messages. |
| Thread category | Categorization/ | Categorization | Certain | Categories: financial |
| | activity/ | Activity | | market, free chat (non- |
| | competence | Competence | | financial), technical |
| | 7 | 1 | | support/faq. Useful to |
| | | | | support competence |
| | | | | analysis |
| Thread number of | Activity/ | Activity | Certain | Number of messages are |
| posts | Competence | | | indicator of discussion |

| | | | | interest |
|-------------------|---------------------------|--|---|--|
| Company Thread | Activity/ Competence | Activity / Competence | Certain | Indicator of how diverse are members interests |
| Poll | | | | |
| Open Poll | Visibility | Activity | Certain | |
| Poll replies | Activity | Activity | Certain | Poll interest indicator |
| Poll category | Competence | Competence | Certain | Finance or else? |
| Post | | | | |
| Write Post | Activity | Activity/Competence | Certain | Basic activity of a member |
| Post Date | Time | Persistency, regularity | Certain | |
| Text length | Activity | Activity | Certain | Indicator of degree of activity. |
| Text Content | Activity/ Competence | Activity, Competence, Disposition | Medium – automatic text analysis not fully certain | Is the message about finance or not? |
| Quote | Activity, Connectivity | Activity, Connectivity, Popularity-Authority | Certain | Quoting a message. Implicit link among messages and respective authors. Sign of interests / social connections |
| Quoted | Competence | Competence/Activity | Certain | |
| Attachment | Activity | Activity | Certain | Attachment usually requires more effort than a post |
| Attachment Type | Competence | Competence | Medium: automatic attachment analysis non fully certain | It is important to discriminate between attachments that are news, company graph or non finance-related images |
| Virtual Portfolio | | | 1735 Members | |
| Rank | Competence | Competence | Certain | Rank based on % of profit |
| Num. Transaction | Activity | Activity | Certain | |
| Profits | Competence | Competence | Certain | % of profit or loss |
| Private Message | | | | |
| Sender/Receiver | Competence | Competence | Certain | Rank on % of profits |
| Number In/Out | Activity | Activity | Certain | |

Data and Indicators Selected for Trust Scheme

 t_p = present time (i.e. time when data were collected, Feb. 2008)

 t_{reg} = time of user registration

 t_{last} = time of last post

 t_{first} = time of first post

 A_{post} = action of writing a post

 N_{post} = number of messages written

 N_{3d} = number of open threads

 N_{att} = number of attachments

N_{poll} =number of open polls

 N_{faq} = number of messages written in the faq/support section of the forum

 L_p = length of a message

 $\mu(Lp)$ = average length of a message

C_{trading} = set of the threads, messages and attachments in the trading section of the forum

 C_{help} = set of the threads, messages and attachments in the help/faq section of the forum

 C_{extra} = set of the threads, messages and attachments in the free chat section of the forum

 N_{news} = number of terms related to news in a message (see Competence Trust scheme for more details)

 $N_{trading}$ = number of trading-related terms in a message

 N_{pdf} = existence of an attachment on line in the message

 A_{quot} = percentage of messages in which a member quotes another member, i.e. a measure of each member's attitude to quote.

Q_{tot} = number of times a user is quoted in his/her messages

 $Q_{a,b}$ = number of times member a quotes member b

 $N_{\text{wa,b}}$ = number of messages member a and member b have in the same threads

 $N_{wa,b,t}$ = number of messages member a and member b have in common in a thread

 N_{thr} = number of threads where the users wrote

 $N_{wp(thr)}$ = number of messages written per user per thread

 P_{info} = presence of personal info

P_{id} = presence of Id skype/msn

After this preliminary analysis we present the matched trust schemes and the critical questions analysis. Our modus operandi is the following: first we present the trust scheme matched and its basic computation, then we go trough the critical questions discussion. The discussion produces:

- a value of plausibility for the scheme that quantifies how plausible the scheme is in the context
- a final computation of the scheme, addressing where possible the critical questions issue. The computation has also a value of plausibility that affects the scheme's plausibility. We largely described how defeaters can be present at the level of the computation and, even if the reason embedded in the scheme is very promising in the context, the impossibility to compute the scheme plausibly makes it useless or potentially harmful
- On the base of the final plausibility value, the scheme can be totally defeated and removed by the computation or just diminished

The critical questions are described in chapter V.

Time-based Trust Schemes

All the information shared has a timestamp, and the time-based scheme can be applied easily.

Critical Questions Analysis (common to all time-bases trust schemes)

Is the environment competitive? There is evidence that the environment is competitive. A preliminary analysis of the distribution of entities' longevity (gap between registration and last activity) shows high variance. This, as described in chapter V, is a first argument supporting the plausibility of the scheme: if the entities had all the same longevity, the scheme would lose selective power.

A forum about trading online is populated by people interested in the stock market. According to a poll, 95% of the entities in the forum trade shares.

Entities disappear from the forum for several reasons that encompass one of the following: bad personal performances, being banned, becoming unpopular or having bad arguments, being an occasional trader. In a poll conducted in the forum, 76% of the entities replied how the longevity of a member indicates his/her competence. The 2000 fall of the market after the .com speculation reduced the number of occasional or even professional traders. On the contrary, in time of

positive market members tend to survive more easily. Surviving collapsing markets is therefore considered a strong evidence of competence. Backed by this analysis, we consider the strength of this trust scheme high.

The above analysis shows our modus operandi: in order to satisfy some critical questions a domain-specific investigation could be required.

Regarding the cost of staying alive in the system – another critical question for time-bases schemes -, there is no cost associated; therefore no supporter can be derived from this CQ.

Regarding the computation used, we focus on each scheme separately.

Longevity

The basic computation determines to simply evaluate:

$$Longevity = t_v - t_{reg}$$
 (6.1)

but this basic computation can be more plausible if we use the time of the first and last activity as the limit of the interval, information that is available in the forum statistics. Therefore:

$$Longevity = t_{last} - t_{first}$$
 (6.2)

By using those two timestamps we implicitly state that the action of writing a post – of any content – is enough to consider the entity active. Following our analysis of chapter VI, section 6.1.2, it should also be considered whether the activities performed at t_{last} or t_{first} were enough to consider the entity operating in the system, or stricter conditions should be defined to consider a timestamp valid. The computation could be enhanced also by considering as valid timestamps only the ones related to messages in the financial-related section of the forum, neglecting the non-financial ones. This could avoid the situation of an entity still active in the system but only providing extra information and having the last message about trading long time ago.

We tested formulas 8.1 and 8.2 expecting better results in the second one, considered more plausible.

In conclusion, considering the meaning of the scheme and the plausibility of the computation, the scheme longevity is considered to have a high plausibility we set to 0.8.

Persistency

Persistency is computed using the formula described in chapter VI section 6.1.2 using the timestamps of the action *writing a post*. Persistency is computed for different values of the constant π as described in table 6.4.

Critical Questions Analysis (specific to Persistency and Regularity)

Are the entities supposed to be persistent?

In stock market trading, there is no reason why an entity should not act without regularity or persistency; long pauses of inactivity are not justified by some environmental constraints and entity activity is not somehow limited or subject to a cost.

Are there cycles in the activity?

Weekly cycles should be considered, since the market is closed two days per week. Therefore it is not request that an entity is persistent during the closing days. We encompassed this in the computation when π was set to one day.

Choice of π

Values of π equal to one day, one week or one month resulted to be effective, showing a good predicting value, i.e. the ability to discriminate entities, that is not related to trust but a required feature for the scheme to be meaningful. In particular, they act as incremental filters: entities showing low persistency with a high value of π are poorly ranked, while entities with high persistency despite a low value of π are the top-ranked.

Time interval
$$(t_{last} - t_{first})$$

Since it is easy to be persistent if the member is a relatively young entity, we considered this time factor in the ranking of the scheme, to avoid a confounding variable where young entities are all top-ranked and old entities are disadvantaged. Entities are divided in tiers according to the size of $(t_{last} - t_{first})$ and a final rank is obtained using a statistical interpolation described in Appendix A, that avoids to consider an entity more persistent than another for factors such as age.

Enhancement of the computation

In the light of the critical question, can we make the computation more plausible? The second computation proposed consider only working days – but also holidays if it is to the advantage of the entity – and, as we did for longevity, we use the time of first and last activity to set the boundaries. All the above discussion stands for regularity as well.

Table 6.4 Persistency Trust Scheme Analysis

| Basic Computation Parameters (Basic formula as described in chapter VI, section 6.1.2) | $\pi =$ time interval equal to 1,7,31 days $A_{post} = \text{action of writing a post OR starting a discussion} \\ T = 1 \\ T_0 = T_{reg} \text{ as starting time} \\ T_{last} \text{ as final time}$ |
|--|---|
| Critical Questions Analysis | Enhanced computation: Considering the working days in the weekly interval Considering $T_{\textit{first}}$ instead of T_{reg} and T_{last} instead of T_{p} as for longevity |
| Plausibility | Set to 0.9 (very high). High for the scheme, high for the computation. Data are complete and certain |
| Uncertainty | None. Data are complete and not affected by uncertainty |

Trust Scheme Stability

No evidence was identified to make this scheme plausible in the context. Members are not defined by complex features. A sign of stability could be the changing of personal information, but too few members update them to be considered a good indicator.

Activity-based Trust Scheme

Members of the forum are supposed to be active by providing analysis, comments, suggestions and various attachments. Indicators of a member activity are:

- 1. Number of messages posted
- 2. Number of threads opened
- 3. Attachments provided
- 4. Number of polls opened
- 5. Number of transactions in the virtual portfolio

A basic computation, as shown in table 6.5, is simply made by aggregating the ranking of the above actions using an average approach, discarding the transaction in the virtual portfolio – only less than 2% of users use the portfolio – and by aggregating the action of opening a thread and a poll, considered at comparable level. The basic computation is therefore:

$$R_{tot} = R(N_{post}) + R(N_{att}) + R(N_{sd} + N_{poll})$$
 (6.3)

Critical Questions Analysis

Cost of activity. Cost free.

Are entity supposed to be active? Yes, inactivity is not regarded as positive evidence in an information sharing community, as largely discussed in chapter IV (see Carter [Car05])

Are there different types of activity? Are there relationships among them?

The basic computation of activity considers the total number of messages posted, the number of attachments and the number of threads opened or polls started.

The critical questions suggest checking if there is a hierarchy among different types of activity, or if some action can be considered compulsory.

For the evidence "number of messages", we consider the length of the messages as well, subtracting the presence of some quoted text in it. Of course a short message could have a good content but, defeasibly, one-word messages cannot be compared to a detailed analysis of the market. The action of posting a message is considered the compulsory one, since it is the basic activity of a forum. A member is supposed to write messages and therefore he can have a high activity only according to the messages written.

The action of attaching files is considered a non-compulsory activity that can increase the rank of an entity but not decrease it. Note how in table 6.5 we use the optional aggregation described in chapter 5 as F_{agg} . The same stands for the action of *opening a thread*, that can increase the rank but not decrease. This allows entities with a high number of attachments or entities that opened several discussions to increase their rank even without a great number of messages posted.

All the indicators are tested in relation to the age of a member to avoid this confounding variable with a procedure analogous to the one used for persistency and described in appendix A.

Corrective coefficient for forum activity

In order to keep our computation more plausible, we checked the trend of the average activity of the forum, in order to discover potential confounding variables.

Our results – appendix A – show how in the early years of the forum the quantity of activity per member was lower of about 15%. Members were used to be less active, while in the most recent years members – even the same members – seem to have more confidence with the forum and write more. The gap is evident in the early years and tends to be smaller in recent years, since 2005 approximately.

In order to take this factor in consideration, we introduce a corrective factor, dependent on the year of the forum, to be applied every time the evidence linked to the activity is used.

Table 6.5 Activity Trust Scheme Analysis

| Basic Computation | $R(N_{post}) + R(N_{att}) + R(N_{3d} + N_{poll})$ |
|-------------------------|---|
| After Critical Question | if (N _{att} < 2) then |
| | if $(R(N_{att}) > R(L(p)*N_{post}))$ then |
| | $B = (R(L(p)* N_{awp}) + R(N_{att}))/2$ |
| | else |
| | $B = R(L(p)* N_{post})$ |
| | end if |
| | end if |
| | if $(R(N_{3d}) > B)$ then |
| | $A = (R(N_{3d}) + B) / 2$ |
| | else |
| | A = B |
| | Rank = A+B |
| Plausibility | High for the scheme, medium/high for the computation. |
| | Set to a value of 0.7 (medium-high) |
| Uncertainty | For younger entities. Data are certain and complete |

Competence/Pertinence

We need to look at computable and plausible signs of competence inside the users' activity and profile. The scheme is highly plausible - posting competent messages is a good sign of trust - but the problem is to understand if it can be computed in a plausible way.

A first easy-to-verify sign of competence is represented by the category a posted message belongs to, as we should divide between the trading and the free-chat sections of the forum. As we checked, it is very rare that a message in the non-financial section, containing free chat, is related to finance.

Regarding the content of the message, signs of competence could be mined by trying to understand if the message under analysis is about trading. We defined a vocabulary of trading-related terms and we searched for their occurrence in the messages. The set of terms comes from an online dictionary of trading downloadable at *FinanzaOnline.it* augmented by a manual analysis of 500 messages to identify recurrent terminology in the community. A list of the vocabulary can be found in the appendix B. This automatic analysis is prone to error, but our task is to estimate its degree of defeasibility. We performed a check over a set of 500 messages (300 related to finance, 200 to others), looking at occurrences of our identified terms, and manually checking if the automatic prediction was right. The result was a degree of accuracy of 78.4% for

trading related messages and of 74.6% for non-trading messages. This accuracy impacts the value of plausibility of the scheme computation.

Regarding the attachments, it makes a clear difference if the attachment under analysis is a graph, a technical analysis, a *pdf* document containing financial news or report, or if it is not related to trading. Graphs are clear indicators of competence, since they are often commented and elaborated to show market trends. Even in the thread dedicated to trading, almost half of the pictures are not related to trading. For each member with more than 500 messages, we performed a sampling of his/her attachments choosing a size that guarantees a confidence level of at least 0.9, dividing the files in the following categories: graph or technical analysis, news and extra.

Table 6.6 Competence Trust Scheme Analysis

| The confidence level required is 90% ($Z_{crit} = 1.68$) | | | | | |
|---|--|--|--|--|--|
| Divided in: | | | | | |
| Graf = number of graph attached | | | | | |
| News = number of news attached | | | | | |
| Others = number of attachments not related to trading | | | | | |
| We defined two vocabularies: | | | | | |
| TermTrading = terms connected to trading | | | | | |
| TermNews = terms connected to news | | | | | |
| We computed: | | | | | |
| Avg_Mex_tr = average number of trading terms per messages | | | | | |
| Mex_tr = number of messages containing at least one trading term | | | | | |
| Avg_Mex_news = average number of news terms per message | | | | | |
| Mex_news = number of messages containing at least one news term | | | | | |
| Num_trafing = Mex_news + Mex_tr | | | | | |
| Num_att_nontrading = number of attachments not related to trading | | | | | |
| min(R ⁻¹ (n_arena/N _{post}),R(num_trading))+ | | | | | |
| R(att_news)+ R(n_att_graf)+ R ⁻¹ (num_att_nontrading)) | | | | | |
| The scheme has high plausibility; the computation, due to the | | | | | |
| automatic parsering of messages, is set to medium (other parts of | | | | | |
| the computation are very plausible). The overall scheme is set to a | | | | | |
| value of 0.75 | | | | | |
| None | | | | | |
| | | | | | |

Computation

The formula used is the following:

$$\begin{split} R_{tot} &= \min \left(R^{-1} \bigg(\frac{n_{arena}}{N_{post}} \bigg), R \bigg(N_{trading} \bigg) \right) + R \bigg(N_{a`news} \bigg) + R \bigg(N_{a`graph} \bigg) \\ &+ R^{-1} \bigg(N_{a'nontradig} \bigg) \quad (6.4) \end{split}$$

Where N_{arena} are the number of messages posted in the *arena* section, the free-chat section of the forum. The pertinence level of a member increases if:

- it shows a high number of messages related to trading or/and graph attached
- it does not write much more on the arena section than in the trading section

Writing in a section of the forum not related to trading does not represent bad evidence as far as the member is providing competent messages in the trading section as well, as the use of the minimum aggregation in formula 8.4 function shows.

All the indicators are compared considering the number of messages and entity's age. The relation with age is important since it turns out that attachments were less common in the early years of the forum as described above.

Social Role Trust Scheme

(Connectivity and) Authority

In an online forum entities are constantly exchanging information, adding new contributions or backing or attacking other members' contributions.

The trust scheme connectivity and authority can be applied by relying on the action of quoting messages as an implicit indicator that the content of the message is somehow considered important. The scheme defeasibly states that well-cited entities are more authoritative and therefore there is a defeasible reason to trust. The scheme is stronger if it is clear when an entity is recognizing the authority and quality of another one, i.e. when the link to B made by A is a clear sign in favour of B's quality.

Critical Questions Analysis

Plausibility of the implicit connection link

Regarding the evidence used, the action of quoting a message can be considered a reasonable evidence that two members are interacting, usually discussing something. This can be strengthened by considering the length and complexity of the message shared; by considering if the two members under consideration had had interactions among different threads, frequently and for a long time. The fact that a message is in the same thread has less importance than quoting a message, but still an evidence of interaction. A thread can contain no more than 500 messages.

The flow of the private messages would have been perfect evidence to spot the relationship among entities, which entities are at the centre of the interaction and which are peripheral, but data, even sanitized, was not accessible. The connection should be divided by forum categories: a

strong connection outside the financial session of the forum is not evidence that the member has some kind of authority for trading issues. The scheme states defeasibly that high connections are evidence for trust, therefore assuming that connections are positive. The action of quoting a message can only reveal that the content of that message is somehow important, but no guarantee is there if the message refers to a positive or negative judgement, or something not related to trading. Therefore the plausibility of the scheme should be decreased and set to medium/high, since our computation does not remove the uncertainty about the *sentiment* included in the quotation.

Another test to be performed is every members attitude to quote. If a member rarely quotes, quoting is stronger evidence that the quoted message has a certain importance.

Testing of the small world hypothesis

In order to test the small world hypothesis, a crucial critical test for social-based schemes, we need to consider the distribution of interaction among members and its degree of connectivity. We consider the following two metrics a sign of interaction between two entities:

- A quoting B (B quoting A counts as a separate interaction)
- A and B writing on the same thread

Our set is composed by 5105 members. The possible interactions are therefore $N^2 - N = 51052 - 5105 = 26,055,920$. The number of interactions of type one are 216,697 (around 0.7%) while the number of interaction of type 2 are 2,312,456 (around 8%). This means that a member quotes in his lifetime around 40 others members and shares discussion with around 450 members (that maybe did not discuss with him directly). If we set a threshold for the minimum number of interactions between A and B in order to consider A and B linked, the figure decreases quickly. Moreover, in order to complete the small world test, we need to check if the length of connections between two random members has a logarithmic complexity. We selected a random sample of 200 couples of members, and using the above metrics to define when an interaction occurred, we collected the average number of connections needed to link the members composing the couple. The average number of connections shows a logarithmic complexity of 5-6 hops to reach two random members. We conclude how the small world hypothesis is verified and members tend to form clusters where they interacts the majority of time, clusters that are linked to other clusters by some *central* entities. The small world hypothesis makes the connectivity scheme more plausible: having connections is more unique and stronger evidence.

Table 6.7 Connectivity Trust Scheme Analysis

| Basic Computation | If $R(Q\%) > R(Q_{tot})$ then | | | | |
|-------------------------|---|--|--|--|--|
| (citation) | $Rank = (R(Q_{tot}) + R(Q\%))/2$ | | | | |
| | Else | | | | |
| 。 第一章 | $Rank = R(Q_{tot})$ | | | | |
| After Critical Question | If N _{awo} > 3 then | | | | |
| | $Rank = Rank + R(N_{qn}) + R(avg(Q_{member}) + R(std(Q_{member}))$ | | | | |
| | For each member, use A _{quot} to weight the number of quoting time | | | | |
| Plausibility | Scheme is high, computation medium. Therefore the level | | | | |
| | of plausibility is set to 0.6 | | | | |
| Uncertainty | Data is certain, message content is uncertain | | | | |

A basic computation simply focuses on the number of times a member's messages are quoted. If a member has a high percentage of quoted messages, this can be optionally used to increase his rank. The percentage of quoted messages for a member is an absolute indicator of how the message of a member has an impact on the community.

In order to make the computation more plausible, we removed again the confounding variable age, number of messages and forum trend. The quantities considered are the following:

$$Q_{tot} = num. of times a member is quoted by all the members = \sum_{b} Q_{a,b}$$
 (6.6)

$$\begin{split} Q_{\%} &= \frac{Q_{tot}}{N_{post}} \; (for \, a \, single \, member) \; \; (6.7) \\ N_{q \, member}(a) &= number \, of \, members \, quoting \, A \; \; (6.8) \end{split}$$

$$Q_{member}(a, b) = number of times member a quoted member b$$
 (6.9)

$$A_{quot} = \%$$
 of times a member quotes in relation to its number of messages (6.10)

If member A quotes B, member A considers the content of B's message somehow interesting. The number of times A quotes B and vice versa will define the degree of connectivity between A and B.

We do not consider outliers with a very low number of messages and we consider the number of members that quoted the member under analysis and the distribution of their messages, to verify if the member under analysis is quoted from a variety of members and not only by few members

several times. This is a sign of connectivity and increases the plausibility of the scheme's computation. We note how we apply the *pluralism* trust scheme to the action of quoting.

The predisposition of each member to quote, represented by A_{quot} , is used to weight the number of Q_{tot} . If A_{quot} is low, a quoted messages has more importance and vice versa.

Info provisioning

This trust scheme refers to information provided by the entity that can be useful for the community. Actions linked to this scheme are:

- 1. Opening a Thread
- 2. Writing messages in the support section of the forum
- 3. Providing recommendations, i.e. votes for the internal reputation system
- 4. Providing information for banning bad users

We note how the last two pieces of information 3 and 4 are not publicly available. Regarding the first evidence, the action of opening a thread is a service that a member provides in respect to the community, since he provides the space for starting or continuing a conversation. Anyway, this piece of evidence alone is not enough to assess the scheme.

The messages a user posts in the section of the forum dedicated to technical assistance, FAQ and suggestions gives clear evidence that the user has a kind of *care* for the forum: he wants to suggest improvements to the forum, report a problem, ask for clarifications. According to Carter, this is an evidence for the information providing trust scheme.

Anyway, too few messages are present in this section in order to build a solid ranking without uncertainty. Finally, evidence 3 and 4 are not public and cannot be collected. Therefore the scheme is not applied since a plausible computation cannot be performed.

Visibility

The scheme provides positive trust evidence for those entities that give personal information that makes them accessible. In our online forum, information available is

- Existence of personal and biographical info
- Contact information such as *skype* ID, *msn* ID, personal URL or blog.

Information can be found in the profile page of each member. The critical questions analysis stresses the problem if the information declared is actually valid and verifiable. It is possible to check the existence of a skype/msn ID or URL. It is hard to understand – and not automatically -

if they are linked to the person under analysis. Therefore, little value is assigned to this information. A manual check can be performed for a subset of members to estimate the likelihood of data to be valid.

Because of this, the plausibility of the scheme is set to low/very low and it is excluded from the computation. In experiment IV we actually check if this assumption is true: we compute the info provisioning trust scheme and we investigate its efficacy.

Past-Outcomes

The past-outcome trust scheme is not used in this experiment, but we perform an analysis in experiment III. Past-outcomes seem to be effective in this scenario, since a member whose advice generates profits or avoids loss has surely a high reputation.

Difficulties arise in computing the scheme, since it is hard to define a plausible way to verify the validity of advice and if a member actually gained or lost.

An ideal computation would be to link members' pieces of advice or predictions to the real trend of the share under analysis and quantify profit or loss. A lot of information can be extracted from a textual analysis of the messages. Some members are keener to give explicit advice, some members post their transactions on the forum to clearly show what they did.

An issue that can make the computation less plausible is the definition of the temporal interval, to be considered in case members did not declare their transaction (buy or sell) with precisions. No one can guarantee the reliability of members' declarations, they may be just trying to gain reputation and visibility. Moreover, it is hard or impossible to automatically implement such computation.

Trust Scheme based on statistic and grouping

We considered this scheme not applicable for the non-feasibility of the computation. The non-applicability of this trust scheme is mainly the fact that members are not enough profiled, few members update it and not much information on them is disclosed to group them and plausibly apply the scheme. We note how we still use a categorization scheme by considering the section of the forum a member is writing in (financial/non-financial/support section) and prejudicing the content of a message on the basis of its category.

Recommendation System

FinanzaOnline.it has an embedded recommendation system – called Popularity level – where users can vote other users, a vote cannot be repeated until a certain amount of votes is given to other users, and votes are slightly weighted according to the popularity of the member voting. In this experiment we do not use the values provided by the internal reputation system for the simple reason that our comparative metrics is based on users' votes, therefore the use of a reputation system would result in a circular reference. Moreover, a preliminary analysis of messages shows how members of the forum consider the reputation value not credible. Experiment II deals with the problem of recommendation system and its validity.

Defeasible Argumentation Phase

Trust schemes are aggregated using the rules described in Chapter 5, table 5.8. Due to the trust scheme used, the rules involve mainly time-based trust scheme, activity and competence. We remind two main argumentation rules:

- Lack of competence cannot be tolerated, even if other trust schemes are positive
- Lack of activity for an old entity cannot be tolerated
- High activity has low value without persistency and regularity

6.2.2.1 Results evaluation

We describe the results obtained by performing different series of computation. In order to evaluate the efficacy of results we need to define efficiency metrics for our trust computation. We propose a first metric E that is defined as follows:

$$E(n) = \frac{1}{n} \sum_{x=1}^{n} C_{rank}(x) - T_{rank}(x)$$
 (6.11)

Where:

- *n* is the number of members included in the metric
- $C_{rank}(x)$ is the ranking of member x according to the community survey
- $T_{rank}(x)$ is the ranking according to our trust computation.

E(n) therefore measures the average error generated by our computation for the set of members considered. For our computation the values of E(10) and E(30) are especially relevant.

Other metrics used are N(10) and N(30), which return the number of top-ten/thirty entities (according to our pool) that are considered in the top-ten/thirty in our computation as well. A low value of E and a high value of E are signs of efficiency.

Table 6.8 and table 6.9 show the results obtained, globally and for each trust scheme. The results show an overall very high degree of precision with an average error of 3.4 positions for the top-ten entities using a set of 5 015 members. We believe this is one of the major contributions of this work: by starting from a set of generic reasons to trust, by investigating their plausibility in the context, and by combining them using an argumentation-based approach we obtained an extremely high precision of predictions.

Table 6.8 Experiment I overall results

| | Before Critical Questions | | | | After Critical Questions | | | | |
|-------------------------|---------------------------|-------|-----------------|------------------|--------------------------|-------|-------|-----------------|------------------|
| Trust Scheme | E(10) | E(30) | Plaus. Comp. | Plaus. Scheme | Defeated? | E(10) | E(30) | Plaus. Comp. | Plaus. Scheme |
| Time-Based | | | | | | | | | |
| Longevity | 278 | 502 | L | Н | N | 112 | 400 | Н | Н |
| Persistency | 102 | 297 | М | VH | N | 31 | 120 | Н | VH |
| Regularity | 161 | 279 | М | VH | N | 48 | 109 | Н | VH |
| Activity-Based | | | | | | | | | |
| Activity | 165 | 392 | М | Н | | 19.6 | 91.8 | VH | Н |
| Competence | 75 | 211 | М | VH | N | 75 | 211 | М | VH |
| Social-Based | | | | | | | | | |
| Authority | 104 | 398.7 | М | Н | | 104 | 398.7 | М | Н |
| Accessibility | 2147 | 1863 | L | VH | Y | | | | |
| Past-Outcomes | | | L | VH | Y | | | | |
| Recommendation | 923 | 1803 | L | VH | Y | | | | |
| Global Aggregation | 112.8 | 171.4 | | | | 4.1 | 65.3 | | |
| Global Argumentation | 88.7 | 143.2 | | | | 3.4 | 40.7 | | |

Table 6.9 Experiment I overall results (2)

| Trust Scheme | E(10) | E(30) | N(10) | N(30) | N(150) | MAX | STD |
|---------------|-------|-------|-------|-------|--------|-----|-------|
| Aggregation | 4.1 | 65.3 | 7 | 13 | 25 | 528 | 103.1 |
| Argumentation | 3.4 | 40.7 | 8 (9) | 20 | 28 | 440 | 93.9 |

The introduction of defeasibility makes the results more efficient. If we consider the computation without defeasibility – i.e. without defeating bad arguments and strengthening good ones – the overall results have, in the best case, an average error of more than 80 positions for E(10) and more than 100 for E(30). The main reason is the usage of two very implausible trust schemes, such as recommendation (see experiment II) and accessibility, that our critical questions invalidate, leading to results 80% better. Schemes such as the time-based and activity-based were very effective individually.

We now comment the results divided by trust scheme.

Time-based Scheme

The 30 most trustworthy entities are all "old" ones. The forum of *FinanzaOnline.it* was opened at the end of 1999, and the youngest of the top 30 entities registered in February 2004 (4 years old), while the average age is about 6.7 out of 8 years of the forum life. 13 entities have more than 7 years. Anyway, many other old entities are not trustworthy, so the scheme has only a one-way validity. We note how, introducing a more plausible computation, values are better than 60%.

Entities are persistent, the top 30 entities' average time of non-interaction is less than one week (5.3 days), and only three entities in the top 30 had an idle time more than two months in their forum life.

Activity

The top 30 entities are very active. They hold the top five position, they usually - but not always - attach files to their messages, confirming the validity of the hypothesis that consider attachments optional. Three of the top 10 entities rarely or never attach a file. On the contrary, these entities show a very high writing activity. Finally, all the top 30 entities usually start conversation, and this makes the difference with ordinary entities.

Competence

The top entities have good signs of competence. Anyway, 5 of the top 30 entities do not have a very high score. These entities show a good number of trading messages, but are also keen to chat and give contributions to other sections of the forum not related to trading. The community –and our computation- does not regard this as bad evidence, as far as they keep

writing messages of high competence as well. We also remind how our computation cannot be considered totally plausible.

Connectivity

The top 30 entities show a variable behaviour in this factor. The top 10 entities perform well and are usually well quoted by a high number of members, but among the top 30, 3 of them have a very poor scoring. Despite of this, the community judged them among the more trustworthy. These entities have a good score in the other factors, but it seems they do not interact with other entities.

Argumentation vs. Aggregation

The results obtained using argumentation rules opposed to a simple aggregation strategy are encouraging. E(30) is now reduced from 65.3 to 40.7, with an error decreased of about 40%. A benefit is achieved for E(10), which is also reduced by 20%.

Using the aggregation strategy, N(10) has a value of 9 (the only entity out of the top ten is 21°), N(30) has the top value of 19 and only one out of the thirty monitored members is out of the first 150. Moreover, the maximum error is reduced and the standard deviation of the error is slightly reduced.

By analysing results in details, we note how aggregation gives to our results more consistency, mainly by reducing the impact of entities with high but not regular activity (it is common to find entities that in one year wrote what normal entities write in 5-6 years and then they disappear from the forum), or by reducing the ranking of entities with high activity but low pertinence, and excluding old but inactive or medium-active entities, that still have a good score in trust scheme not related to activity, such as longevity, pertinence, authority.

We also remind how an argumentation is implicitly performed at the level of trust scheme critical questions, and results show huge improvements of about 80% between the application of the schemes with or without defeasibility.

We also note how the first three more trustworthy entities are the same for the two computations, meaning that very trustworthy entities are recognizable even with simple and untested evidence. On the other hand, entities without such a strong consensus, argumentation adds value, as the increment in the value of E(30) shows. More precisely, we performed a paired *t-test* between the aggregation and the argumentation ranking of the top-30 entities in order to investigate the

statistical significance of the two sets of results. The degree of freedom of the paired *t-test* is 29, giving a critical value t of 1.65 and 1.97 for 90% and 95% confidence level respectively. The t-test was verified with a value of t of more than 3.4 for E(30) and with a value of 2.33 for E(10).

6.2.3 Exp. II - Full application of the method using a limited set of evidence

In this experiment we perform a trust computation by using only the information that is immediately visible and available to any forum user and we compare the efficiency of this computation with the one of Exp. I. Aim of the experiment is to investigate and quantify if the complexity introduced by our method (a broader set of evidence, plausibility study and argumentation) produced an added value that justifies its application. Second aim is to investigate if the introduction of defeasibility and argumentation over this limited subset improves the results of a computation based on this restricted set of evidence, in order to have a further evidence to support our hyp. 1 and 2. The data *FinanzaOnline.it* makes visible to its users (members and unregistered guests) are the following:

- 1. Date of registration
- 2. Number of messages
- 3. Frequency of messages
- 4. Last activity
- 5. Value of the Reputation System

For instance, figure 6.3 shows some of the data publicly available to any visitor of the forum: is this information enough to take decision about trust?



Figure 6.3 Some visible data of a member: Date of registration, Total Messages, Popularity Value

We note how the first two indicators are linked to time, the third is related to activity and the last one is linked to social role / reputation. Therefore, the information sustains a basic application of our method performed over a subset of evidence. Note how evidence 1, 4 were

used in experiment I to compute longevity, evidence 3 represents a kind of regularity or persistency factor, the number of messages (evidence 2) was used in the basic computation of activity, while the internal reputation system was not used.

The following table shows the results of the application of the four groups of evidence (the last activity is used to re-compute longevity as we did in Exp. I).

Table 6.10 Experiment II overall results (1)

| Evidence used | E(10) | E(30) |
|--|-------|-------|
| Date of Registration and Last Activity | 112 | 400 |
| Reputation | 923 | 1803 |
| Frequency | 273 | 493 |
| Number of Messages | 214 | 343 |

As shown in the table above, only *longevity* is a quite effective scheme – if we use the registration date and the last activity date as in experiment I - but frequency leads to results of little value in comparison to our trust mechanism persistency and regularity.

The reputation systems appear strongly unreliable, with only two of the top 30 entities in the first 30 positions and none of the top 10 in the first ten positions. The reputation system is therefore ineffective in assessing trustworthiness as already anticipated in our preliminary investigation.

A more quantitative way of showing the lack of plausibility of the internal reputation system is to perform the positive bias test. A positive bias is evident if reputation values tends to be high for the majority of users, losing predictive power and ability to discriminate bad and good entities. Motivation behind positive bias is a situation where entities do not want to harm each other for fear of revenge or a situation of collusions.

The positive bias of FinanzaOnlie.it is evident if we consider the distribution of the members of the forum according to their reputation value divided in 100 intervals. If we consider all the members, the system seems to show some selective distribution of values (figure 6.4), but with a suspicious peak of high values and the absence of intermediate values.

If we consider only entities with at least 400 messages – around 10 000 members – i.e. we exclude inactive entities, the situation is the one depicted in figure 6.5. The reputation system has lost any predictive power. If 85.8% of entities have the maximum reputation value, how can the reputation value help us to select the most trustworthy entities?

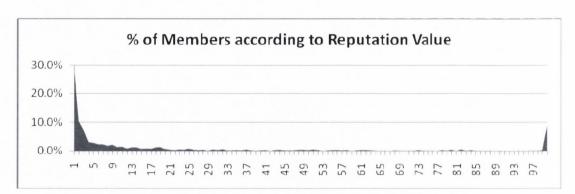


Figure 6.4 Percentage of Members according to their Reputation Value

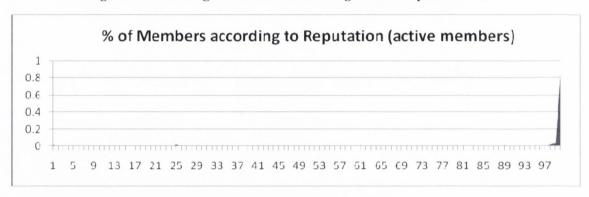


Figure 6.5 Percentage of Members according to their Reputation Value. Only active members considered

Table 6.11 shows the final aggregated results. The simple aggregation strategy gives invalid results, with a value of E(10) and E(30) of 261 and 439, due mainly to the negative effect of recommendation and the lack of efficiency of the rest of the computation.

Results can be directly comparable with our computation. Considering just the number of messages, representing the best evidence, only 3 out of the top 30 members are correctly predicted, and only 2 of the top 10. The average error, for the top 10 positions, is 166, with a maximum of 673 and 423 for two entities, and with 5 top-10 entities out of the first 300 positions, meaning that these five members are considered normal entities in the forum. If we consider the top 30 positions, there are 9 members above the 1000° position. We note how in our computation the worst entity among the top 10 was placed in 35° position, and the worst of the top 30 in 440° position. We now wonder if these results can be improved using defeasibility.

Argumentation with the basic formula

Can we improve the result of the basic computation using defeasibility?

By using the evidence that is visible to the user, we try to improve the results of the simple aggregation strategy in order to investigate again if argumentation is effective. Using similar arguments that inspired the critical questions and the argumentation layer of our method, we could start investigating the plausibility of the evidence used.

The internal recommendation system is not a plausible evidence for all the reasons described above. We should give it a lower impact or completely defeat it.

The longevity is a plausible factor, and there is no uncertainty in the data. The same stands for frequency.

The number of messages is again a plausible factor, since they represent the only evidence of interaction of an entity in the forum.

Moreover, we should apply our argumentation layer, that is mainly reduced to the relationships involving activity, longevity, frequency (considered a simplified regularity scheme) and reputation. The results obtained are again shown in table 6.11.

Table 6.11 Experiment II: Argumentation vs. Aggregation

| | E(10) | E(30) |
|---|-------|-------|
| Aggregation Final Value | 261 | 439 |
| Argumentation and Reasoning Final value | 148 | 288 |

There is a strong increment in the quality of predictions. Results are still not efficient on an absolute scale, but the application of the method reduced E(10) by almost 50% and E(30) by almost 40%.

Final comment

We performed a trust computation based on a set of evidence directly visible to the user. We aggregated them with a simple averaging strategy and using defeasibility.

The results obtained are the following:

- The computation is not effective, the evidence used show little validity. In particular, only the total number of messages show some degree of validity
- The aggregation strategy based on averaging does not produce better final results
- The introduction of the combination of critical questions and argumentation improves the computation by 40%-50%. An increment of 50% is obtained for the top 10 entities.

Therefore we can conclude the following:

- The extra complexity introduced by our method, that investigates a broader set of
 evidence and implies a series of tests over this evidence adds value to the computation
 efficiency
- Evidence aggregated using a defeasible argumentation layer is more effective

6.2.4 Exp. III - The Defeasibility of Past-Outcome Trust Scheme

In this experiment we investigate the outcome-based trust scheme with the aim of showing its defeasible nature. We remind, as discussed in the previous chapter, how this trust scheme is the most popular in trust computation but only in [Fal04] and few other works it has been critically analysed, while it is usually used straightforward and regarded as the most objective trust mechanism. This experiment shows how this objectivity can lead to bad trust computations if not properly tested and how a critical question analysis avoids non-negligible mistakes.

In the context of *FinanzaOnline.it*, the past-outcome trust scheme has a clear meaning: a positive outcome is the gain traders accumulated on the stock market. The main problem is removing the uncertainty factors. In order to apply the mechanism, we should know the trading transactions of a member, information that is clearly private and not available.

In an online forum there is an extra-dimension, represented by the trading advice that a member gives the community. Of course, the two pieces of information are not always linked: a member may gain money while giving false advice on the forum.

The second information is important in order to assess the trustworthiness of a member in a public forum: what builds his/her reputation are the pieces of advice/predictions he makes publicly on the forum. Of course, a member may not post explicit suggestions, but only opinions or analyses. Due to the difficulties of collecting reliable information without a manual time-consuming check of messages posted, we discarded the use of this scheme in experiment I.

In this experiment we investigate the scheme by performing this analysis on a subset of forum members. We isolated and used as evidence only situations in which the forum member expressed a clear suggestion. It is impossible to understand if the suggestion was malicious or not – maybe the member wanted to harm somebody else with a bad suggestion, but this information is not necessary, as far as this information is public and contributes to his/her reputation value.

Therefore, the most trustworthy entities are the one that give advice on shares that increase their value and avoid shares that decrease.

We performed a textual analysis of trading messages looking for trading advice to be linked to their authors. We selected only clear advice, preferring short or medium term predictions than long term one, harder to verify. Both suggestions about entry point (i.e. sell) or exit level (i.e. buy) had to be clearly checked, and fuzzy or unclear suggestions discarded.

Nevertheless, some uncertainty remains. Given a trading suggestion, it is not possible to know if this was first-hand information of an entity or it was taken by another entity.

An answer is, again, that when members post suggestions, they take the responsibility of what they write in front of the other readers; even if it was not their own advice, they are backing it.

In many cases, however, it was possible to map suggestions to their owners by following the flow of messages.

Furthermore, by focusing mainly on short-time trading messages it is easier to match them with their owners, since the flow of messages is limited usually to a single day/week and one or two forum threads.

In the experiment we selected a sampling of trading advice stratified as follows:

- A set of messages containing 60 suggestions coming from the top-10 members according to the forum community
- A set of messages containing 60 suggestions coming from members between the 100° and 1000° positions of the trust ranking (average trust level)
- A set of messages containing 60 suggestions coming from members above the 1000° position (low level of trust)

Suggestions were equally collected over the entire life of the forum, from 2000 to 2008 and divided by 2-year periods. Suggestions were also divided according to a risk variable: high, medium and low risk companies. The risk level associated to each company is a piece of information taken directly from the forum (risk levels associated to companies are a common information in financial market analysis).

We then matched the suggestions with the actual trend of the suggested company, using past data series. By doing so, we could count the number of successful suggestions or failures.

The experiment is described in the following tables 6.12-6.14.

Table 6.12 Experiment III: Messages Collected

| Number of Messages analyzed | | | | | |
|-----------------------------|-------|-------|-------|-----|-----|
| 12 - 30 | 02/03 | 04/05 | 06/07 | 08 | |
| Top 30 | 61 | 89 | 112 | 90 | 352 |
| Average | 67 | 118 | 152 | 88 | 425 |
| Low | 111 | 166 | 301 | 105 | 683 |

Table 6.13 Experiment III: Percentage of Good Suggestion

| Good Suggestion (%) | | | | | |
|---------------------|-------|-------|-------|----|---------|
| | 02/03 | 04/05 | 06/07 | 08 | Average |
| Top 30 | 42 | 63 | 77 | 40 | 55.5 |
| Average | 39 | 65 | 75 | 32 | 52.75 |
| Low | 34 | 58 | 66 | 26 | 46 |

Table 6.14 Experiment III: Good Suggestions divided by Risk Level

| Good Suggestion per risk | | | |
|--------------------------|-----|--------|------|
| | low | medium | high |
| Top 30 | 18 | 19 | 13 |
| Average | 17 | 18 | 15 |
| Low | 16 | 17 | 17 |

Analysis of results

Table 8.16 shows the general results of the past-outcome scheme. The top-30 entities have an average of good past outcomes of 55.5%, while the second-tier entity an average of 52.75%, and the low trusted entity of 46%. There is a statistical significance between the top-30 and the last tier, but not between the first two tiers. The scheme shows some validity, but its extent is smaller than trust schemes such as time-based or activity-based.

This first observation shows how it is possible (defeasibly) to predict users' trustworthiness with an analysis of the quantity and time-distribution of their activity without knowing the outcomes of the interactions.

Second, the table clearly shows the defeasible nature of the mechanisms. For instance, if we consider only the period of time between 02/03 and 06/07, the first and second tier entities have exactly the same number of good suggestions, while during the period 04/05 the second tier entities have the maximum score, leading to contradictory results. It is easy also to appreciate the fact that during periods such as 06/07, the percentage of good suggestions is very high (more than

75%) while during the period 02/03 is less than 40%. As one of our critical questions attached to this scheme suggests, the reasons for this results are the external market constraints in which the entities had to operate. In years of bullish market – such as 2004-2006, it was easy to give good advice, even for non-expert traders, while during bearish market – such as 2001-2002, suggestions were harder to give and the most expert entities made the difference. This shows the importance, when evaluating past-outcomes, to consider the *non-controllable market conditions* in which the entities are interacting.

Moreover, we noticed other important factors: the top-30 entities are keener to give suggestions, as shown in tables 6.12 and 6.13: in order to collect 60 suggestions, we analysed only 352 messages for tier 1, while the number is double for tier 3 entities (682). We noticed also how the number of messages analysed is closer in periods of bullish market and the gap is wider in periods of bearish market (see 2006 and 2008 results). Our proposed interpretation is the following: what makes the difference is the ability to provide advice: trustworthy entities provide more suggestions, even in turbulent times, risking their reputation and showing a better control of the situation. On the contrary, average traders do not expose themselves during periods of bearish market. We could consider this as an instance of visibility and accessibility trust scheme, since trustworthy entities proved to be accessible and visible in hard times as well.

6.2.5 Exp. IV - Info Provisioning Trust Scheme Analysis

In this experiment we performed an analysis of the information provisioning trust scheme. For a group of selected entities, we checked the presence of information such as *msn* or *skype* contact, personal web page, personal data (age, profession, hobbies) on their forum profile page. As already described, this trust scheme is not considered plausible in the context by our critical questions: a minority of members fills this information and it is hard or impossible to check if the information is genuine. This is why the potential values produced by the trust scheme were not introduced in Exp. I. In this experiment we investigate if this presumption is correct, or if it would have been better to use the scheme, neglecting our plausibility analysis.

Information such as website or msn or skype contact can be partially tested, even if it is not possible to link the information with the real identity of the member.

We performed an experiment on a sample of members, including the top 30 members, to see if the presence of information, without investigating their validity, is a sign of trust. The experiment verifies hypothesis 2 if the scheme results not effective.

For each member, we trace the presence of biographical information, personal information (such as profession and hobby), contact information (skype, msn, email, blog) and the number of visits on the personal web-page, provided by the forum and used as an indicator. The sample of members is composed by 150 entities divided in 3 groups: the top 30 entities, 30 entities chosen from 30° to 200° position, 30 entities chosen from 200° to 5000° position. The results are shown in table 6.15.

Table 6.15 Experiment IV: Argumentation vs. Aggregation

| | Bio | Personal Info | Contacts | Visits |
|-------|-------|------------------|----------|--------|
| TOP30 | 67.7% | 22.6% | 28.6% | 392.6 |
| AVG | 65.6% | 16.3% | 25.0% | 414.8 |
| LOW | 57.7% | 19.2% | 28.9% | 448.8 |

By looking at the results obtained, it is easy to verify how there is little or no statistical difference among the three groups of data. If we consider personal information, entities in the lower tier show almost the same level of contact information provided as the top-30 entities. The scheme therefore is not an effective indicator of trustworthy entities, as predicted by its plausibility analysis. Our hypothesis 2 is therefore verified.

6.3 The case of Wikipedia: a trust reasoning for wiki-based applications

Wikipedia is a global online encyclopaedia, entirely written collaboratively by an open community of users. The Wikipedia project started on the 15th of January 2000 [Wik06c], and it is now composed by a community of more than 1.5 million registered users, delivering 2.5 million articles in the English version, and millions of articles in almost 200 different languages. It is one of the most 10 visited web-sites on the web, and it represents the most successful and criticized example of collective knowledge, a concept that is often lauded as the next step toward truth in online media.

The set of experiments described here have the aim to investigate the trustworthiness of the two crucial entities composing the domains articles and users.

This experiments face an increasingly acute problem in today's digital world. The problem of identifying trustworthy information on the World Wide Web is becoming increasingly acute as new tools such as wikis and blogs simplify and democratize publications.

Wikipedia represents the most extraordinary example of this issue, but its quality has recently been questioned. On one hand, recent exceptional cases have brought to the attention the matter of Wikipedia trustworthiness. In an article published on the 29th November 2005 in USA Today [Wik06a], Seigenthaler, a former administrative assistant to Robert Kennedy, wrote about his anguish after learning about a false Wikipedia entry that listed him as having been briefly suspected of involvement in the assassination of both John Kennedy and Robert Kennedy. The 78-year-old Seigenthaler got Wikipedia founder Jimmy Wales to delete the defamatory information in October. Unfortunately, that was four months after the original posting. The news was a further proof that Wikipedia has no accountability and no place in the world of serious information gathering [Wik06a]. On the other hand, Wikipedia is being discussed not only negatively. In December 2005, a detailed analysis carried out by the magazine Nature [Gal05] compared the accuracy of Wikipedia against the Encyclopaedia Britannica. Nature identified a set of 42 articles, covering a broad range of scientific disciplines, and sent them to relevant experts for peer review. The results are encouraging: the investigation suggests that Britannica's advantage may not be great, at least when it comes to science entries. The difference in accuracy was not particularly great: the average science entry in Wikipedia contained around four inaccuracies; Britannica, about three. Reviewers also found many factual errors, omissions or misleading statements: 162 and 123 in Wikipedia and Britannica respectively. Moreover, Nature has stated that, among their scientific collaborators, 70% of them had heard of Wikipedia, 17% of those consult it on a weekly basis and about 10% help to update it.

In the first set of experiments, we apply our defeasibility-based trust computation to the problem of article trustworthiness. The independent trust rank used for our comparative evaluation is the internal quality system of Wikipedia, whose strong plausibility is investigated in the next section. Then, section 8.3.2 describes the phases of our method, including the representation of the domain, the mapping of the trust scheme and the evidence collection, the defeasible computation performed using the matched trust schemes.

Again, the three key hypotheses will be evaluated: the effectiveness of defeasibility, the impact of argumentation and the efficiency of the method on an absolute scale.

In particular, some trust schemes not used in the previous set of experiments – notably stability and similarity-based - are here investigated.

The second set of experiments investigates the trustworthiness of Wikipedia authors, the registered members of the community. The comparative evaluation is now based on an internal Wikipedia award system, called the *Barnstar* award, whose plausibility is also investigated in the next section. The aim of this second set of experiments is not another complete application of the method. We rather focus on the time-based trust scheme, representing one of the major novelties that this thesis introduced in trust computation.

Finally, the third experiment shows how the authority trust scheme, applied in an implausible way, leads to poor results. Therefore, the experiment is an evidence for hypothesis 2. The discussion of these experiments is shorter than the FinanzaOnline.it one, omitting some procedures that have been exhaustively described in the first experiment.

Dataset used and plausibility of the comparative evaluation

The dataset used in experiment V – Wikipedia articles analysis - is composed of 7718 articles, including all 846 featured articles plus the most visited pages with at least 25 edits. The dataset was retrieved from Wikimedia [Wik06] on the 17th March 2006. These articles represented the majority of the editing activity of Wikipedia (around 61% of all edits), thus it can be considered a significant set.

Regarding the validity of the internal quality system, we consider this mechanism highly reliable, as the already cited Nature's experiment [Gal02] proved. The internal Wikipedia quality system is a reputation system managed by the community itself. Every article can enter the process of nomination, and it has to pass a strict procedure involving peer-review, check of standard and guidelines, completeness of the topic, absence of contradictions and editing battle and so forth.

There are two levels of award: the lowest is the good article status: articles contain excellent content but are unlikely in their current state to become featured; they may be too short, or about a too extensive or specific topic, or on a topic about which not much is known. The featured article status, the highest award, is the target of our evaluation. It represents the excellence of Wikipedia: less than 0.1% of articles has the featured article status.

Regarding experiment VI, analysis of Wikipedia authors, our set of data is composed by about 5000 authors, including 325 awarded authors, all of whom wrote in the section "English literature", randomly chosen among the most developed sections of the encyclopaedia. The award selected as comparative metric is the *Barnstar Award*, assigned using an internal mechanism of nominations and voting by Wikipedia authors. According to Wikipedia guidelines, this award is given "as a reward for those users who make outstanding contributions in the particular field." [Bar08.]. We consider this award a trust evidence plausible enough to perform an interesting investigation. A strict minority of users is awarded and the motivations are always very detailed, public and compatible with the human notion of trust. The main grounds are the reliability of the authors, the quality of their editing, their constant care for the community.

Relying on the *Barnstar Award* is still plausible, but less plausible than the mechanisms used for featured articles. It is less strict - a single user can assign it, even if it is rare -, it is less visible than an award given to articles, harder and unlikely to retract (while featured status can be retracted many times) and less efficient. More information can be found in [Bar08].

6.3.1 Exp. V - Trustworthiness of Wikipedia Articles

Wikipedia Application Model

The application model used is depicted in figure 6.6. Our Wikipedia model is composed of two principal objects (wiki article and wiki user) and a number of supporting objects. Since each user has a personal page, the objects *users* and *articles* share common methods such as creating, modifying and deleting edits or uploading images. An article contains the main article page (class wiki page) and the talk page, where users can add comments and judgments on the article. Wiki page includes properties such as length, a count of the number of sections, images, external links, notes and references. Each page has a history page, containing a complete list of all modifications. A modification contains information of User, date and time, the text of the old version. The community of users can modify articles or add discussion on the article topic (the talk page for that article).

Evidence Selection, trust scheme, critical question

Following the same procedure described for the *FinanzaOnline.it* experiment, we matched trust scheme to available elements of the domain. Wikipedia allows us to match several trust schemes due to the quantity of public information it makes available, consequence of its

open philosophy. We now describe the list of applicable trust schemes, with details of the pieces of evidence used and the computation performed.

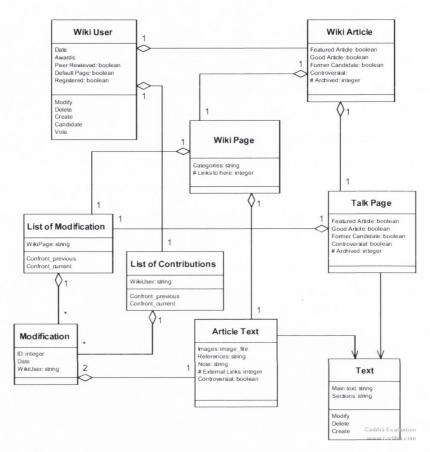


Figure 6.6 – Wikipedia Application Model

Activity

The trust scheme suggests considering the degree of activity of an entity as a defeasible reason for its trustworthiness. In the context of Wikipedia, signs of articles activity are essentially two: the article is read and edited. Activity can be mapped onto the following actions:

- Number of visits
- Number/analysis of edits
- Number of comments on the talk page

A more sophisticated computation, following an analogous discussion performed in the *FinanzaOnline* experiment, is the one that considers the different types of activity. Having edits is considered the essential activity, since it is essential for article's existence.

The number of comments on the talk page reveals if article's context is undergoing a review from Wikipedia community members, since it usually contains suggestions or critique regarding the content of an article. Finally, the number of visits defeasibly suggests that an article is more controlled than an article that is seldom or never accessed: the article content is read and consumed many times by occasional readers or Wikipedia members, and potential errors can be spotted more quickly. Seigenthaler's article was edited only 10 times during the four months, much less than the average for that type of articles.

This last evidence is therefore linked to a sense of controllability of an article. We could say that in this particular context the trust scheme activity is in the middle between activity and accessibility and transparency.

A starting computation of the scheme ranks the set of articles according to the above parameters. Therefore, the basic computation is the following:

$$R_{tot} = R(edits) + R(talk)_{optional} + R(visits)_{optional}$$
 (6.12)

The defeasible tests introduced in chapter VI focus on the different type of activity that may be performed. The action editing can have variable complexity, and a better analysis of this complexity should be introduced. An editing could be measured by its length or attachments, an approach that could be criticized but better than assigning to all the edits equal status. Moreover, we remind that the activity trust scheme is focused on the quantity.

Second, a critical question suggests that the size of the entity should be considered: a small entity with high degree of activity may still result less active then a large entity with low activity. The size of an article is in Wikipedia the *importance* of its topic. It is obvious that articles such as "United States of America" register higher activity than an article about the village of Abbeylinx in the centre of Ireland. Here the attacking argument is that we are not measuring the degree of activity but only the importance of an article.

These observations make the first mapping of the scheme implausible. In order to address this problem, we may refine the computation as follows. We consider the relative size of an editing, measured by its length, removing the reverted edits and distinguishing between attachments and minor or major editing.

We also consider the importance of an article as a weight to judge its degree of activity. In order to estimate the importance of an article, we suggest using two indicators: number of visits – removed from the computation of activity and now used for checking its plausibility - and number of Wikipedia links that point to the article page, therefore using a kind of *PageRank*

approach to assess the article's importance (what *PageRank* is supposed to do) – and not its trustworthiness!

Stability

In the context of Wikipedia, stability of an article is defeasible evidence that the article has reached a mature and complete status, the content is accepted and it is not source of contentions and editing revisions.

Critical Question Analysis

Regarding our critical questions, stability seems plausible in the context of Wikipedia. First, assessing stability makes sense because entities are subject to change. In a wiki-based application documents are constantly edited and under revision. Nevertheless, in the context of Wikipedia an accepted content should remain stable, while an article that needs to be expanded, or is not in a complete status, is likely to be subject to change.

The plausibility increases by considering the threshold for stability, i.e. which is the magnitude of the change in order to consider the article still stable. This calls for an analysis of the editing, and a way of discarding minor or negligible edits.

Wikipedia authors usually mark their edits with the prefix minor or major, and any automatic change executed by automated procedures is also marked. An analysis of the edits, as we perform for the trust scheme activity, can increase the plausibility of the computation.

Moreover, it is also likely that an article may be subject to change only in its form and not in its content. A detailed analysis should consider the content of the changing.

All these reasons makes stability a strong argument, but not completely feasible to verify, and therefore we set the plausibility of a value of 0.6.

Regarding the critical questions inter-schemes, stability does not mean anything without activity. A stable article despite a high number of edits and with a high number of visits is plausible trust evidence. We can defeasibly presume that the article has a good content if it is constantly revised and refined. On the contrary, an article that is never edited and visited is an abandoned article, rather than complete and reliable; stability is no longer a piece of evidence. Thanks to our argumentation layer, this situation will be defeated.

Computation

Regarding the computation to be performed, Wikipedia provides all the past version of an article making the computation feasible. We define the function

$$N(a,t):t \to n \tag{6.13}$$

That gives the number of edits done at time t for a specific article a. Then we define:

$$E(t) = \sum_{t=0}^{t_p} N(a, t)$$
 (6.14)

that, given a timestamp t it gives the number of edits done from tine t to the present time t_p for a specific article. We than define

$$Txt(t): t \to L$$
 (6.15)

that gives the number of different words between the version at time t and the current one.

We evaluate the stability of an article looking at the values of these two functions. If an article is stable it means that E(t), from a certain point of time t, should decrease or be almost a constant that means that the number of editing is stable or decreasing: the article is not being to be modified or disputed. The meaning of Txt(t) is an estimation of how different was the version at time t compared to the current version. When t is close to the current time point, Txt goes to θ , and it is obviously θ when t is the current time. An article is stable if Txt, from a certain point of time t not very close to the current time is almost a constant value. This means that the text is almost the same in that period of time. The computation performed, due to the complete data available, does not reduce the plausibility of 0.6 already defined in our previous discussion.

Pluralism

Pluralism plays a key-role in any wiki-based application. As an article may be edited by any author, it is important that it reflects an unbiased point of view emerging from a multitude of editors. Pluralism of an article can be computed by looking at the sequence of editing performed by the various authors, considering each contribution relative importance. Moreover, the dynamic of the *talk page* such as the number of authors leaving comments could also be important.

Due to its importance in a wiki-based application, as internal Wikipedia guidelines confirm, pluralism has a very high plausibility of 0.9. We do not set plausibility to 1 since it is not possible to define the relative importance of editing with high accuracy without a manual analysis of it.

Computation

In order to compute the degree of plausibility of an article, we apply some of the basic operators described in chapter V over the distribution of numbers of editing per authors.

In particular, we consider the

- Average number of edits per authors
- Standard deviation of number of edits
- The contribution of the top 5% C(5%) and the contribution of the last 20% 1-C(80%). The operator C is described in chapter V.
- The contribution of the authors with more than n edits per article, useful to discard authors with very few edits

A more plausible computation is performed by giving a weight to the edits based on its length and considering also the edits of the talk page in the same way.

The computation is plausible thanks to the availability of the historical information, and the value of plausibility of the scheme is confirmed to be at 0.8

Longevity

Longevity and other time-based trust scheme are not used. They are not considered plausible. Article's longevity, i.e. its age, can be obtained by Wikipedia without uncertainty. Nevertheless, articles' creation seems more linked to article importance, and affected by the fact that articles are often merged or re-arranged into other articles. Therefore, longevity is not used as a trust argument, but still computed since it is used by other trust schemes critical questions.

Regarding persistency and regularity, they make more sense for authors' activity, and we prefer not to consider them in the computation.

Similarity/Categorization

The trust scheme categorization suggests that an entity should exhibit properties in line with its category, and outliers' entity should deserve further investigations. Outliers are suspicious, therefore they do not gain trust easily but they do not deserve distrust either, except for the case where outliers can be plausible considered an incomplete or immature entity.

In Wikipedia the trust scheme Categorization can be easily mapped to the embedded classification of Wikipedia articles.

In order to compute the scheme, we need to define a set of article's properties used to describe how the standard looks like.

We performed a computation using a basic list of properties of articles (such as length, number of images, number of edits..), and we selected the most important features as the results of a principal components analysis as described in [Don07a] and reported in Appendix C.

The plausibility of the computation is high, since Wikipedia categorization system is reliable and very detailed. Anyway, a category may contain different articles with different features because of their topic, and our presumption that an average exists cannot be sustained. The plausibility of the scheme is therefore questionable and considered defeated.

In order to increase the plausibility of the scheme, we also consider the importance of an article based on the linking structure as described above.

Therefore, we produced two computations, one using the category system of Wikipedia, second considering also the relative importance of articles. Articles of comparable importance in the same category should have similar properties. Entities are ranked on the basis of their distance from the average of the category they belong to, as assigned by Wikipedia. Since this average is a sampling of the actual average – we do not have all the articles – an uncertainty is attached to the computation as described in chapter V. If the uncertainty is more than 0.9 the argument is defeated and not used in the computation for that specific category.

The compactness of each category, given by the standard deviation of properties distribution, is used to modify the strength of the results: outliers in compact distributions are more severe and vice versa. The plausibility of the scheme is increased by these observations, and it is therefore set to a medium value of 0.6.

Standard Compliance

The trust scheme standard compliance suggests that an entity is trusted to the extent of its degree of similarity to a standard. The plausibility of this scheme depends greatly on the existence of an accepted standard and its applications.

Wikipedia provides a dedicated section of the encyclopaedia where guidelines are provided about how a Wikipedia articles should look like. Guidelines are provided the outlook of articles, articles template, length, layout, balance of editing, sections, notes, linking structure, references and so on. The complete list of guidelines and polices is available here [Wik08b]. For a subset of

guidelines, mainly comprising features linked to layout and editing (length, template presence, references, sections, we performed an automatic analysis of our articles, ranking them according to the presence/absence of these features. Considering critical tests and the available knowledge of the domain, we set the trust scheme's plausibility to 1 (highest value), since there is a strong consensus among Wikipedia community around the set of guidelines, and it is almost certain that an article that wants to claim the status of *featured* must adhere to such standard in order to be considered so. Due to the partial nature of our computation (not all the guidelines are controlled), the general plausibility is set to 0.8.

Authorship

The value of authorship used is the output of the following experiment, where we compute a value of trustworthiness for each Wikipedia authors.

This value is used to compute articles trustworthiness by averaging all the trust value of their authors, using the amount of editing done by each author in the article as a weight. Plausibility is set to 0.7, encompassing a high value of the scheme but hard to compute, as discussed in the next experiment.

6.3.1.1 Results

Table 6.16 show the results obtained. The metrics used are the following:

- *C*: correlation between the distribution of standard and featured articles according to their trust value. If *C* is low, featured and standard articles have been efficiently divided.
- %FA: percentage of feature articles that in our computation have a trust value greater than 70%
- %SA: percentage of standard article that in our computation have a trust value greater than 70%

Note how the metrics are based on the normalized values of each scheme, following a different strategy than the ranking-based used for FinanzaOnline.it experiment.

The results obtained are first analysed for each trust scheme.

Stability. Stability produces a good value of C (32.4) and a rate of good featured article of 64%. Overall results are of average value compared to the other schemes, but still valid. %SA is 45,

which is quite high, but this value is reduced by the defeasible argumentation that cuts stable but inactive articles.

Table 6.16 – Experiment V overall results

| Eseller | Ве | efore | Critic | al Ques | stions | | | Afte | r Critic | al Questi | ons |
|-----------------------|------|-------|--------|-----------------|------------------------|-----------|------|------|----------|-----------------|------------------|
| Trust Scheme | С | %FA | %SA | Plaus. Comp. | Plausibility Scheme | Defeated? | С | %FA | %SA | Plaus. Comp. | Plaus. Scheme |
| Time-Based | | | | | | | | | | | |
| Stability | 32.4 | 64 | 45 | 1 | 0.7 | N | 32.4 | 64 | 45 | 1 | 0.7 |
| | | | | | | | | | | | |
| Activity-Based | | | | | | | | | | | |
| Activity | 32.5 | 60.1 | 34 | 0.8 | 0.7 | N | 27.4 | 69.3 | 26.5 | 0.9 | 0.8 |
| Pluralism | 22.3 | 77.4 | 16.1 | 1 | 0.8 | N | 14.2 | 83.2 | 11 | 0.9 | 0.9 |
| 医大型等 二、例如 | | | | | | | | | | | |
| Grouping | | | | | | | | | | | |
| Similarity | 61 | 69.1 | 61 | 0.3 | 0.6 | Υ | 41 | 57.8 | 57.8 | 0.6 | 0.6 |
| Standard- | 19.1 | 82.6 | 14.9 | 0.8 | 1 | N | 19.1 | 82.6 | 14.9 | 0.8 | 1 |
| Compliance | | | | | | | | | | | |
| | | | | | | | | | | | |
| Social-based | | | | | | | | | 7 | | |
| Authorship | 41.5 | 65 | 53.2 | 0.6 | 0.8 | N | 39.5 | 67 | 53.2 | 0.6 | 0.8 |
| 4. 关于1. 证书 | | | | | | | | | | | |
| Aggregation | 30.2 | 62 | 22 | | | | 21 | 77.9 | 16.4 | | |
| Argumentation | 25.1 | 68.8 | 19 | | | | 18.4 | 80.8 | 13 | | |

Activity: compared to stability, activity basic computation shows similar results for C (30.5) and a slightly worst predictive power (60.4). The computation after our tests decreases C (27.4) and increases FA to 69.3, with an increment of almost 10% (15% relatively). Therefore, we verified our hypothesis that our modified computation should have been more plausible and effective.

Pluralism. The computation was considered very plausible (0.9), and the scheme produces a very good value of C (22.4) and %FA 77.4.

Similarity/Categorization. The scheme produces unclear results. C has the highest value of 61% meaning that featured and standard articles have not been divided. FA% has a value of 69%, which is very positive but %SA has also a value of 61%, meaning that the scheme is not effective in discarding standard articles and it considers them of good quality. Featured articles are rarely outliers, but also many standard ones are usually not. The scheme therefore is not completely effective. Our second computation that considers the importance of an article and considers the

compactness of the distribution shows better – but still the worst - results, with C down to 44%, a similar value for %FA and %SA reduced of 3.2% down to 57.8%.

Standard Compliance. The scheme produces much better results than similarity and very good results on an absolute scale: a value of C of 19.1%, a value of %FA of 82.6%, and a value of 14.9% for SA. This means that featured article are compliant to a standard while standard articles are not.

Authorship. Using our trust values computed in experiment VI, we obtained a value of 41.5% for C and 65% for FA%. SA% is 53.2%, meaning that trustworthy authors write on featured articles but on standard article as well, and therefore their presence is not a very effective way to discriminate articles. In line with our method, we should investigate the reason why the computation is not effective as predicted. A defeating argument not identified could be present. A candidate could be the fact we had not a way to identify the leading authors and assign to such authors a higher importance.

Argumentation vs. aggregation and overall results

As table 6.16 shows, argumentation among evidence increases the quality of the combined results. In both cases, argumentation increases the quality of about 15%. On an absolute scale, the results obtained are very positive. In order to better present the overall results, we analyse the graph in figure 6.7, published in [Don07a].

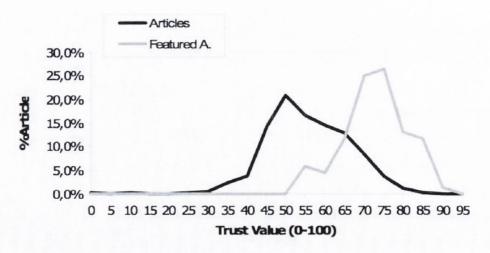


Figure 6.7 – Featured and Standard Articles distribution

Table 6.17 – Experiment V overall results (2)

| Correlation | 18.4 % | | | | |
|-----------------------|---------|-------------|--------|--|--|
| | % of FA | % of SA | GAP | | |
| Bad: TV < 50 | 0 | 42.3% | 42.3% | | |
| Average: 50 < TV < 70 | 22.2 % | 54.7 % | 32.5% | | |
| Good: TV > 70 | 80.8 % | 13 % | 64.8 % | | |
| Very Good: TV > 85 | 13.2 % | 23 articles | 13.2 % | | |

Figure 6.7 represents the distribution of articles on the base of their trust values. We have isolated the featured articles (grey line) from standard articles (black line): if our calculation is valid, featured articles should show higher trust values than standard articles. Results obtained are positive and encouraging: the graph clearly shows the difference between standard articles distribution, mainly around a trust value of 45-50%, and featured articles distribution, around 75%.

Among the featured articles, 77.8% are distributed in the region with trust values > 70%, meaning that they are all considered good articles, while only 13% of standard articles are considered good. Furthermore, 42.3% of standard articles are distributed in the region with trust values < 50%, where there are no featured articles, demonstrating the selection operated by the computation. Only 23 standard articles are in the region >85%, where there are 93 featured ones, despite the fact that standard articles are ten times the number of featured ones. The experiment, covering articles from different categories, shows that the method has a promising general validity and a good degree of precision.

6.3.2 Exp. VI - Trustworthiness of Wikipedia Authors

The aim of this experiment is to compute a trust value for each Wikipedia Author. In order to perform our comparative evaluation, we selected as an independent reliable trust metric the internal awards system of Wikipedia, which assigns to the most valuable authors the "Barnard Star Award". Nominees are expression of the community and decided by the community, in a recommendation-system fashion.

Our trust metric is accurate only if it is able to provide awarded users with a high trust value. Note how users without award are not by default untrustworthy, but a difference should be appreciated.

As Wikipedia is an extraordinary example of collaborative and information sharing application, trust schemes derived from Carter's model should be effective. Schemes such as regularity, persistency and information provisioning, if computationally applicable, should be considered highly plausible. Other schemes, such as activity, are also discussed. As usual, aggregation and argumentation will be compared, mainly to check the activity and time-based interrelationships.

Analysis of the Schemes used

Longevity

Longevity is easy to compute without uncertainty. We propose the two computations used in the *FinanzaOnline.it* experiment. Regarding the value of plausibility, we note how the scheme is weaker than in the case of *FinanzaOnline.it*: environment is not selective, authors are not pushed to change their identity if they did something "wrong" – as it is the case for the trading community – and community members consider less the age of a member as an evidence of trust and experience.

Nevertheless, longevity is still evidence that an author has experience and confidence with Wikipedia, and therefore the scheme is not defeated completely, but we diminish its plausibility to low (0.4).

Persistency and Regularity

The scheme has high plausibility in Wikipedia: authors that constantly edit, control and refine articles show their care and commitment for the cause of the Encyclopaedia. Defeasibly, the schemes link these reasons to trust.

Regarding the computation to be performed, all the data are available and certain but, following the same discussion performed in the previous experiment, there is not a fully plausible way of understanding the importance of an editing used to trigger an activity. Nevertheless, this will impact more the activity trust scheme, since persistency and regularity look more to the time of the interaction rather than the amount of activity performed.

The computation is performed using the formula described in chapter VI, using any edits as a sign of activity. Plausibility is therefore confirmed to be high and set at 0.9.

Activity

Activity is computed relying on the action of editing, which encompasses uploading of images as well. Activity is a condition *sine qua non* in order to be considered a valuable contributor to Wikipedia. As written above regarding articles, the computation of the scheme cannot be completely plausible, mainly because of the impossibility to clearly understand contributions' complexity. Therefore, the plausibility of the scheme, very high *per se*, is reduced to 0.8 due to the computation.

Info provisioning

Contrarily to *FinanzaOnline.it* experiment, the scheme in the context of Wikipedia is considered plausible. We remind how the scheme refers to the fact that an entity should provide useful information to the community to fulfil one of its principal social roles.

There are many ways a Wikipedia member can provide information:

- Being an administrator
- Being a watchdog i.e. controlling a list of articles against vandalism and bad editing
- Writing in the section dedicated to Wikipedia guidelines and policies
- Providing discussion on the Talk Page
- Voting and proposing featured article status
- Peer-reviewing articles
- Create templates
- Create Articles

All this information can be fetched with no uncertainty by mining articles' history and personal page of users. The plausibility of the computation and of the scheme is therefore solid, and we assign a value of 0.9.

Accessibility

Contrarily to *FinanzaOnline.it* experiment, the scheme in the context of Wikipedia is considered plausible. Users keep a personal page with personal information covering

biographical, subject of interests, skills, degree and specializations, articles created, written or controlled, template created (already used in the info provisioning scheme), contacts links, the articles to do, the discussion in which the user took part. The presence of this information adds transparency regarding the author.

The usual problem linked to the validity and impossibility to check information is still present, reducing the plausibility of the scheme.

Therefore, the data available support a plausible computation but, since information is not completely verifiable, we assign a plausibility of 0.7

Past-Outcomes

This is due mainly to the absence of explicit judgments about authors' activity. One of this judgements is actually the award the author gained, that cannot be used since are part of the comparative set of data. An idea, not investigated in this thesis but left for future experimentation, could be looking for implicit sign of good/bad outcomes. An example could be the fact that an edit was reverted, cancelled or if it is still present in the current version of the article.

Results

Table 6.18 – Experiment VI overall results

| Trust Scheme | Award | Award | Non award | Non | Plaus. | Plaus | Plaus. | Defeated? |
|-----------------------|-------|-------|-----------|-------|----------|--------|--------|------------|
| | high | low | high | award | Comp.(*) | Scheme | 2.5 | |
| | | | | low | | | | |
| Time-Based | | | | | | | | |
| Longevity | 53.5 | 46.5 | 43.7 | 56.3 | VH | L | 0.4 | N |
| Pertinence | 75.6 | 24.4 | 21.1 | 78.9 | Н | VH | 0.9 | N |
| Regularity | 71.8 | 28.2 | 23 | 77 | Н | VH | 0.9 | |
| | | | | | | | | |
| Activity-Based | | | | | 244 | | | 7 . 7G |
| Activity | 67.5 | 32.5 | 31.2 | 68.8 | Н | Н | 0.8 | N |
| Accessibility | 73.9 | 26.1 | 41.7 | 58.3 | М | VH | 0.7 | N |
| Info- | 77.7 | 22.3 | 27.7 | 72.3 | M/H | М | 0.9 | N |
| Provisioning | | | | | | | | 2 - 20 1 1 |
| | | | | | | | | |
| Aggregation | 72.2 | 27.8 | 29 | 71 | | | | |
| Argumentation | 74.85 | 25.15 | 26.2 | 73.8 | | | | |

^(*) VH=0.9, H=0.8, M=0.5, M/H=0.65, L=0.3

The results are displayed in table 6.18. Metrics used are the following:

- 1. Award-high (true positive): percentage of awarded users that have been considered trustworthy by our computation
- 2. Award-low (false positive): awarded users considered untrustworthy
- 3. Not-award-High (false negative): non-awarded users considered trustworthy
- 4. Not-award-Low (true negative): non-awarded users considered untrustworthy

The percentage for each category is calculated by applying our computation over the set of users, and by considering the 70% value as the threshold for considering an author trustworthy, following the analogous procedure of experiment V.

We remind how, for the specific nature of the *Barnard Star award*, it is important to have good values for metric one and two, while metric 3 and 4 are interesting but not completely plausible. Author's award is less efficient than the featured articles status: awarded authors are plausibly good authors, but many good authors may have not been yet considered for this award.

Longevity

The scheme shows a slight degree of validity since the majority of awarded authors are considered trustworthy (53.5%) but also 43.7% of not-awarded authors have an high score.

The results of the computation are already in the most plausible situation. The computation performed using the present time as entity's last activity is invalid.

Persistency and Regularity

The scheme is very effective: 75.6% of true positive and only 21.1% of false negative. Regularity is slightly inferior then Persistency but still very effective,

Activity

Activity has a good degree of validity with 67.5% of good predictions and 31.2% of false negative

Accessibility

Accessibility resulted to be a very effective scheme, slightly below persistency and despite the plausibility value of 0.7 we assigned. Authors with awards have detailed and up-to-date personal pages and tend to display personal information, showing care and loyalty to the Wikipedia community. On the contrary, it seems that many non-awarded users display their information as well, since 41.7% of them are considered trustworthy. Therefore, in Wikipedia users seem to provide information and being visible, in accordance with the spirit of the

application. The consequence is that this scheme loses predictive power and it became less effective in dividing the two groups of users.

Info Provisioning

The scheme is the most effective: awarded Wikipedians are constant provider of various source of information valuable for the community, as proved by a value of 77.7% of true positive predictions. Regarding not-award users, predictions are slightly worst but still effective.

Aggregation vs. Argumentation

The argumentation layer is reduced to few rules involving longevity, activity and persistency/regularity. The results obtained are better from 3% to 10%.

Regarding the overall results, they show a good degree of accuracy with a rate of 74.85% of true positive and 73.8% of false negative, even if lower than experiment I and IV. Note how the value of the scheme Information Provisioning was better than the final argued value (about 2% better in true positive), the process of combining evidence did not bring any added value in comparison to a single trust scheme. Finally, more than 24% of non-awarded authors are considered trustworthy. As explained above, these set contains good authors still not considered for an award and the data is hard to be interpreted in positive or negative. It could be interesting to notice how many of this authors will receive an award in the future.

6.3.3 Exp. VII - Evaluation of a PageRank-based heuristic for Wikipedia Articles

In this experiment we evaluate a *PageRank*-based heuristic for Wikipedia articles described by McGuiness et al. in [McG06]. The authors propose a metric, based on an adaptation of the PageRank algorithm, to compute a trust value for articles. The metric proposed was the following

$$T_{doc} = \frac{occ([[d]])}{occ([[d]]) + occ(d)}$$

$$(6.16)$$

The meaning of the factors is the following:

- T_{doc} is the trust value associated with a document
- d is the title of the articles, such as "Beer" or "Theory of Relativity"
- Occ(d) is the number of occurrences of the article title d in the whole Wikipedia
- Occ([d]) is the occurrence of the article's title as a link to the article d in the whole
 Wikipedia

Therefore, the metric suggests that there is a reason to trust an article if it appears as a link, meaning that other Wikipedia editors considered that the article should be linked and therefore has a good content. The background reason and the computational implementation is a clear reference to the *PageRank* algorithm, where the linking structure is used to build a metric that many authors, such as Massa [Mas03], proved to be a trust metric (defeasible in our opinion).

In chapter VI we placed this kind of heuristics under the connectivity or authority trust scheme, arguing about the defeasibility of the mechanism when, for instance, the underlying assumption that a link between A and B is an implicit judgment on B's quality cannot be sustained.

Anyway, in Wikipedia, articles are linked for many reasons. First, automatic procedures are in place for linking articles, that periodically scan articles' text and create links for the most important, i.e. visited, articles. This argument creates a first undercutter argument for the metric, by proposing alternative explanations for the action of linking.

Second, if an author, editing article A, creates a link to article B, the following reasons are plausible:

- 1. The author wants A to be complete, so a link is added independently from the content of but only to make A complete.
- If an author wants to create a link to B, there is no choice but a single article. Therefore linking article B does not mean that it is considered better than other potential candidates

 as it may be the case for websites since other candidates simply do not exit.
- More visited and important articles, independently from their content, has more probability of being linked by authors, since it is likely that they know the existence of the target article.

These observations, derived from the critical questions described in chapter VI, are enough to consider the metric of low plausibility. Therefore, according to our defeasible paradigm, it is better to defeat the argument generated by the metric or consider it reduced.

We applied the metric using our dataset of articles of exp. I, using the same evaluation metrics.

Table 6.19 – Experiment VII overall results

| | C(featured/standard) | %FA > 0.7 | %SA<0.5 |
|---------------------|----------------------|-----------|---------|
| McGuinness's Metric | 67.1 | 49.3 | 52.1 |

An analysis of the results obtained show how correlation has a value of 67.1%, much higher than our value (16.2% in the best situation, 30% in the worst case) and even the predictive power of the metric is limited or even erroneous: only 49.3% of the featured articles have a normalized trust value more than 0.7 (about 80%) in our case, and 52.1% of standard articles have a value less than 0.5 against a value of 68.9% in our computation. The metric therefore is not able to selectively divide featured and standard article, and, if applied, would negatively affect a trust computation.

Conclusions

In this chapter we evaluated our trust model with a series of experiments in different domains. Using a comparative evaluation approach, we tested the efficiency of our computation in the prediction of entity trustworthiness. The results obtained show a precision around 75-80% according to the experiment considered. We tested every single aspect of the method, showing how the critical questions paradigm improves the efficacy of the computation and the final argumentation layer increases the results further and gives a better understanding of special cases. We show how the method can provide justifications for its results. Finally, we discussed several issues related to the application of the method in the light of the experiments performed: the class of suitable applications, some analogies with other computer applications, the open issues of mapping dependency, showing what our method can provide to face this problem. We conclude showing how the method it represents one of the few evidence selection strategy in trust, that, even if not definitive and comprehensive, is driven by a generic expertise of trust, that does not delegate the gathering of evidence to domain experts and that provides a way to decompose the problem into isolate problems not related to trust. Our evidence selection strategy is similar to few approach such as Carter's, and has more computational evidence than other high-level models of trust. We stressed also the dialectical scenario in which our method could be used, where each party is responsible for the accomplishment and the validity of the stages of the method, from data gathering to trust scheme mapping and plausibility study, reconciling the method with distributed agents' scenarios.

Finally, we stressed the introduction of defeasibility as the key to extend trust computation using those mechanisms considered deductively too weak to be employed but backed by social science.

Chapter 7 Conclusions

Introduction

In this final chapter we present our conclusions, stressing future works and open issues. In the first section we restate the objectives of the thesis and we provide subsections which summarise how each objective was achieved. In the second section we present the main contribution of the thesis and in the final section 3 we discuss future work and open issues.

7.1 Thesis' Objectives

Our work was motivated by observing the defeasible nature of trust, and by hypothesizing that the introduction of defeasible reasoning techniques in trust computation would deliver positive effects. Our initial research question was the following:

If trust is a form of defeasible reasoning, the computational techniques that researchers in AI and Argumentation Theory have developed for this kind of reasoning could be effective in trust computation.

The research hypothesis raised four issues:

- 1. The validity of the starting assumption trust is a form of defeasible reasoning
- 2. How to use defeasible reasoning and presumptive reasoning techniques in trust, i.e. how to define/design a computable model of trust structured over these two disciplines.

- 3. To implement the designed model
- 4. To define and measure the positive (if any) impacts of the envisaged model The following subsections discuss how the four above issues were accomplished.

7.1.1 The validity of modelling Trust as a form of defeasible reasoning

Our state-of-the-art review described in chapter 2 - the Romano definition of trust - and our discussion in chapter 4, section 4.1, sustained our first hypothesis that trust can be intended, and therefore computed, as a result of a defeasible reasoning.

Trust has largely been seen by social scientists as being both a result of an unconscious or intuitive process and as a result of an evaluation based on evidence and reasoning.

Moreover, we have shown how this reasoning process is defeasible, since the mechanisms composing the reasoning are non-monotonic assumptions and both because they are taken under situations of imperfect knowledge.

7.1.2 Designing a trust model based on Defeasible Reasoning

Our trust model design is presented in chapter 4, while chapter 5 describes our list of trust schemes. Recalling from our introduction, the design of our model dealt to deal with four issues, depicted in figure 7.1. We now describe how we dealt with them.

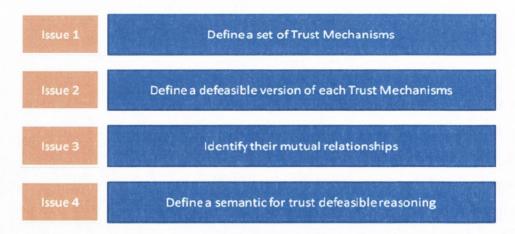


Figure 7.1 Trust Model design's issues

Issue 1: list of trust schemes

We defined a list of trust schemes by investigating the state-of-the-art in chapter 2 and 3. We took into consideration both social studies about trust and both their computational implementations. This list does not claim to be comprehensive but it is a set large enough to reason meaningfully about trust. We provided a list of 25 trust schemes, and for each of them we provided a computational implementation. The list of trust scheme is reproduced in tables 5.1 to 5.7.

Issue 2: Introduction of defeasibility

We dealt with this issue by investigating, for each trust scheme identified, the assumptions on which it is based, in order to identify which elements can be used to test the validity of such assumptions. Central to this task was the analysis of the different incarnation of the same computational trust mechanisms and the study of Castelfranchi and Falcone [Cas02], as discussed in chapter 2. We produced a list of critical questions to be used to assess its plausibility (see chapter 5). The critical question paradigm was modelled after Walton's defeasible reasoning described in chapter 3.

Issue 3: Mutual relationships among the trust schemes

We provided a set of mutual relationships among trust schemes to be used by our defeasible semantic. This set of relationships, provided in table 5.8, have to be intended as defeasible relationships. Relationships are implemented as a set of horn rules with a specific value of plausibility attached.

Issue 4: Semantic definition

We specialized Pollock's semantic in order to compute the status of each argument in the reasoning process. The specialization consists in a new *function of conclusion* that better considers our initial evidence-based computation; an extension of the *accrual function* to encompass not only the *maximum strength* argument, but also the *minimum strength*, the linear combination, the averaging, the optional and compulsory aggregation. These functions let us better fit domain-specific situations where trust has to be assessed. Finally, we defined linear support and attack functions, introducing the concept of supporting function that can increase the plausibility of a reasoning link, absent in Pollock.

General considerations, strengths and weaknesses

Each of our solutions to these issues exhibit strengths and weaknesses that we briefly summarize here. A more extensive discussion is also provided in section 7.4 later in this chapter.

Regarding the list of trust schemes defined, we contend that this list is a valuable tool for performing a trust analysis of target domain applications, as our evaluation has supported. The set is large enough to support meaningful computation and, to the best knowledge of the author, it encompasses the large majority of the current landscape of trust models.

For time-based, activity-based and social-based trust schemes, we provided extensive computational versions. For the group of trust schemes focused on similarity and cognition we still need to fully investigate them from a computational point of view and an effective computational version is not provided yet.

Regarding the critical question paradigm, the list provided is comprehensive enough to test trust scheme validity, but we recognize how some questions have a qualitative nature that hardly can be translated into a quantitative assessment. A clear methodology for their use is still at embryonic stage.

Regarding issue 3, the definition of mutual relationships can be considered a well-developed part of our design. They represent a novelty in the landscape of computational models of trust.

Finally, some aspects of the semantic proposed are still open issues. Its main strength is the effectiveness of its outputs (see evaluation chapter) and the ability to produce results statistically different from a simple aggregation strategy (see chapter 5). Since it has the properties of Pollock's semantic, its theoretical validity is guaranteed by Pollock's conclusions [Pol02]. The meaning of the output values produced has a bias towards trustworthy entities. In fact, the semantic is effective in producing a fine-grained rank for the trustworthy entities, while it just recognises an entity as untrustworthy, without further ranking.

7.1.3 The implementation of the model

In order to implement the model, we used a ranking statistics approach, where the strength of a trust scheme for each entity is proportional to the ranking of the entity in the community. In chapter 5 we described the main advantages and the drawbacks of this choice. In order to

implement the model, we use the semantic and the trust scheme computations defined in chapter 5 over a dataset of entities' activity.

7.1.4 Evaluating the impact of defeasibility

We performed our evaluation using a comparative approach, testing our trust metric with an independent and accepted metric. The metric is independent since its results cannot be used by our computation and it is computed used a separate process, avoiding any circular reference. It is accepted by the community, in order for the results to be useful and of practical utility. As described in our introduction, our evaluation sought to verify three hypotheses depicted in figure 7.2 and discussed in the next subsections.

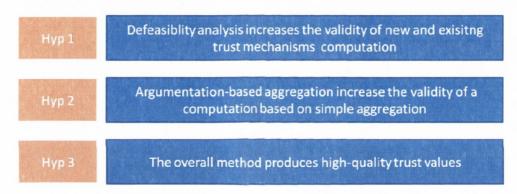


Figure 7.2 Hypotheses to be evaluated.

Hypothesis 1

The introduction of defeasibility positively impacts the trust computation by giving more importance to plausible mechanisms and by invalidating implausible ones. This increases the true *positive* predictions –since plausible arguments count more - and reduces the *true negative* ones – since implausible arguments are defeated.

We tested this hypothesis both in FinanzaOnline.it and in Wikipedia. Referring to our evaluation chapter, experiment 1 showed how the introduction of defeasibility during the analysis of trust schemes increases the quality of predictions of more than 80%. Experiment II showed an increment of 40% while experiment V, in the context of Wikipedia, showed an increment of 35% in the quality of predictions.

Experiment III and IV are also evidence of the validity of our hypothesis, since the trust schemes analysed were correctly predicted to lead to invalid results.

In general, the critical question pattern increases the efficiency of the computation, in some cases drastically. We note how critical questions are derived from the trust scheme, not the domain. On the other hand, the application of trust scheme without testing can lead to contradictory results. The experiment on *FinanzaOnline.it* shows how even the outcome-based scheme, regarded as the ultimate example of objectivity, could lead to poor results when not tested.

Hypothesis 2

More precisely, we expect that a final trust metric based on a defeasible reasoning semantic is more efficient than a metric generated by a simple aggregation strategy.

Experiments I,II,V,VI tested hypothesis 2. After having matched and tested our trust schemes, we compared the results obtained by a simple averaging strategy to the one produced by our defeasible reasoning semantic, as described in chapter 5.

The results are summarized in table 7.1. The introduction of defeasibility has in all our experiments a positive impact, up to 42%. The impact is stronger when many mutual relationships among the trust schemes are used and contradictions or inconsistency are present. Experiment I showed also how defeasibility do not create a huge increment of performance for the most trustworthy users, since the strong consensus among the trustworthiness of these entities do not usually contain contradictions. Our experiments clearly prove also the following: a single trust scheme, even tested, could not guarantee the successful application of the method, largely increased by the use of set of trust scheme.

We also recognise that the hypothesis 2 is the least tested, since the mutual relationships used (and therefore tested) in all our study-cases still represent a limited subset of all possible cases. For instance, as table 7.1 shows, experiment V and VI, where a small subset of mutual relationship were activated, the gap with a simple aggregation is limited.

| Table 7.1 Defeasible semantic versus Simple aggregation | | | | | | |
|---|---------------------------------|------------------------------|------|--|--|--|
| Experiment | Defeasible Semantic | Aggregation | Gap | | | |
| I - FinanzaOnline.it, Full application of the method | 3.4 (average position error) | 4.1 (average position error) | 18% | | | |
| II – FinanzaOnline.it, Limited set of Evidence | 148 (average position error) | 261 (average position error) | 42% | | | |
| V – Wikipedia Articles | 80.8 (% of good predictions) | 77.9 (% of good predictions) | 1.3% | | | |
| VI – Wikipedia Authors | 73.8 (% of good predictions) | 71 (% of good predictions) | 3.8% | | | |

Hypothesis 3

Our third hypothesis was that the overall method would provide efficient trust values. Not only does defeasible reasoning have to increase the efficiency of the trust value (hypothesis 1 and 2) but we have to show our defeasible computation is efficient.

The hypothesis was tested in experiment I and V. The experiment I, performed over FinanzaOnlie.it, showed an error of 3.4 positions out of more than 5 000 members. This result can be considered, on an absolute scale to be high accuracy. Moreover, 9 out of 10 of the most trusted member resulted in the first ten positions.

Experiment V, in the context of Wikipedia, showed an accuracy of predictions around 81% with a rate of 13% of false positive. On an absolute scale, the result can be considered encouraging and of good accuracy.

7.2 Contributions

This section describes the contributions of our work. The main contribution to the state-of-the-art is the definition of a novel computational model of trust based on defeasible reasoning and argumentation. Moreover, we provided an implementation of our model that allowed us to evaluate the impact of our novel trust computation giving evidence of its accuracy and efficacy.

Our contribution is supported by several novelties that our model brings to the computational trust state-of-the-art. These can be summarized into theoretical and computational aspects.

Contributions to the theoretical modelling of trust

- Introduction of Argumentation and Defeasible Reasoning. Our model introduces the idea of trust as an argumentation between two parties, the trustier and the trustee. This dynamics could be implemented in multi-agents distributed systems. To the best knowledge of the authors, argumentation-based trust models are not still investigated. Our contribution is not only in the idea of using argumentation theory in trust, but also providing a list of trust scheme to support such process.
- Definition of Trust schemes and Critical questions. We defined an extensive list of trust schemes. The majority of the schemes are existing trust mechanisms extracted from literature, while some represent computational novelty. Our contribution for the first set

of schemes is the definition of their defeasible version, i.e. the study of the assumptions they rely on. For the second set of trust schemes, we introduce a computational version of them.

- A semantic for trust reasoning. We contributed both to trust model definitions and to Argumentation Theory by defining a defeasible reasoning semantic for trust.
- Evidence selection. Our model is a step in the definition of a general methodology for evidence selection, and a consequent investigation of a new set and source of evidence. By starting from a representation of the domain, our model identifies elements that could be used as trust evidence. The key issue is that our trust model contains a generic trust expertise that justifies the evidence selection. Instead of relying entirely on domain-specific expertise, we assumed that trust is an expertise per se that by its own is able to sustain a valid evidence selection. Domain-specific expertise has a supportive role instead of delivering the solution. Moreover, the generic trust expertise is able to sustain a methodology, giving that systematicity that subjective ad-hoc heuristics approach defects.

Contributions of the implemented model to the way trust is computed

- Introduction of defeasibility. The introduction of defeasible computational techniques and how we adapted them to trust has produced a more efficient defeasible aggregation strategy than the previous ones. This is shown in our evaluation chapter 6 by comparing simple aggregation strategy, usually used in trust computation, to our defeasible computation Moreover, we have obtained a more efficient use of existing trust mechanisms due to the introduction of their defeasibility study.
- We introduced new mechanisms to compute trust, namely temporal factors, activity-based and stability-based mechanisms, as better described in chapter 5, that we proved to be efficient.
- We produced a trust computation alternative to the usual two mechanisms, recommendations and past-outcomes. The two mechanisms, often considered as objective were critically analysed showing their defeasible nature and potential inefficiency. Our computation provides an alternative metric to be used to check the validity of these two schemes, or extra data to be merged in a more efficient decision-making process.

We conclude this contribution sections by underlying some general features of the method that, even though they do not represent an explicit intended contribution, represent interesting and singular examples in the current literature.

Our model is an example of model where the human notion of trust is central. We avoided any reductionist approach but, on the other hand, we wanted to produce a computable model. Rich models, such as the cognitive ones or the original Marsh's models are often criticized for being only high-level pointers to computations that are too demanding or not feasible. Even if our model does not claim to make complex cognitive models computable, it is an effort in the direction of computing trust as a distinctive expertise, a complex phenomenon that calls for justifications and motivations. Our model conclusions is fully justified and the fact that the reasons for the computations are kept explicit and human-understandable; make it a transparent decision support tool.

7.3 Discussion of the method and open issues

In this section we discuss several aspects of the method in the light of the experiments performed. We discuss not only the quantitative efficiency of the method, but also all the issues related to method stages: the analysis of the domain, the collection of evidence and the mapping of trust schemes, the critical questions process, the computation and the aggregation. We also analyses the kind of applications used in the evaluation in order to better define a class of application where our model can be effective. Part of this final discussion is dedicated to future works.

7.3.1 Analysis of the Applications used

We focus now on the type of applications used underlying some common characteristics.

The first feature is the <u>availability of data</u>. In all applications there was a rich dataset describing the domain history and interactions. This setting calls for centralized applications with an accessible database. In fact the applications used are in the context of online web communities with such a database. The set of applications covers many interesting web 2.0 phenomena such as wikis, blog, forums, social networking site. This observation seems to rule out distributed application where data, by nature are incomplete.

A first answer to this issue is that the model must be able to handle the uncertainty coming from unperfected knowledge of data. Our model can handle this, since high uncertain data ultimately make the trust scheme under analysis not plausible, not for its relative strength in the domain, but for insufficient elements supporting the conclusions.

A second argument considers a distributed scenario where trust is the result of an argumentation between a trustee and a trustier, scenario that is the original setting of a presumptive argumentation. In this scenario, like in a trial, it is the responsibility of the entity proposing an argument, using a certain trust scheme, to provide the adequate evidence to make it sustainable. The trustee may use an argument that the trustier may refuse just for lack of information available to sustain the thesis. The trustee has therefore to collect more relevant data to satisfy the counterpart. Data availability becomes therefore part of the argumentative process and it can be used to attack or defend arguments. In conclusion, the problem of data availability in distributed scenario becomes an added defeasible argument in the discussion. This discussion has as an obvious hypothesis: data can be certified to be genuine, a security-related problem that is highly pertinent but not part of our problem space.

The second feature of the applications used is the fact that entities <u>interact with frequency</u>. It is out of discussion that entities that are completely new in the system, or that have very few interactions are the most difficult to assess. Trust scheme like regularity, persistency and activity are largely linked to frequent interactions. Our model can still say something on new entities or entities that are acting with low frequency using mainly trust schemes like transitivity – that links the entity to other entities that becomes a kind of guarantors –, its social connections that he might have despite its not frequent interactions or its competence, that are not based on the amount of interactions but on the qualitative aspects. This dependency on time and activity in our model is a reflection of the human notion of trust and common to many computational trust models, from recommendation-based to outcome-based. Even if the problem of an initial trust value could be solved in a satisfactory way, for a correct trust computation interactions and knowledge are needed to reduce the uncertainty of the conclusions and produce a robust ranking. Finally a trust-based decision on entities not interacting frequently has to be taken considering the available evidence coupled by exogenous factor such as risk and disposition.

A third feature is the <u>collaborative nature</u> of our applications. Domains where there is no interactions between entities are obviously not relevant to trust, nor therefore to our method. Moreover, an extra requirement about collaboration is that the presence of a <u>common goal</u> among

the entities is needed. The absence of a common goal makes our trust analysis useless. Moreover, the common goal must be compatible with a human notion of trust, i.e. being a *rational* one.

This hypothesis is not so restrictive. Several online communities of practices, e-marketplace, any service-oriented scenarios and many self-organizing systems can usually satisfy the hypothesis. Communities where there is no consensus about values, norms and purposes – a typical example being community sharing subjective tastes – make our method less effective.

Classes of Applications

In the light of the experiments performed, the class of application that fits those requirements are:

- Wiki-based and collaborative repository of information, self-organizing system of information.
- Online-forum, community of practice, social-network online community.
- Large database of user activities, connected to a common scope or aim.

The list covers several classical applications of computational trust with some differences.

This class of applications seems large enough to encompass several applications related to the emerging collaborative nature of the Web, including Web-Services. These are perhaps ideal applications since the service-oriented activity of such systems guarantee a certain common scope and a plausible metric for reliability and trustworthiness.

The method also suits more classical data-mining of large database describing community of users linked by a purposeful activity and sharing common aims. Useful applications could be in the enterprise environments, market analysis, and automatic content-quality. Finally, the monitoring aspect of our method can suit scenarios where entities interacting are autonomic agents or distributed devices. Examples are distributed autonomic network of sensors or any self-organizing systems where components interact together. By monitoring the system and applying our method, interesting inferences and indicators about the behaviour (reliability, trustworthiness, and its health status) can be derived.

7.3.2 Similarities with other computer science applications modus operandi.

In the light of the type of computation performed and the input used in our evaluation, we could link our model to other classical computer science applications.

Our implementation is a tool for performing a *data-mining* process that identifies and extracts *trust relationships*. This represents an original application in computational trust. The method has similarities with other field of computer science.

The method can be defined as an *expert system for trust*. Trust is the generic expertise that the expert system delivers. The knowledge base is represented by our trust schemes and their critical questions that contain also the rules to be used to infer new knowledge.

In a pattern-matching fashion, our trust rules are matched on the available facts, represented by the elements of the domain.

By building our trust model, we did the job of the *knowledge engineer*, while the role of the expert of the domain was represented by the social science literature about trust. In applying the model to a specific application, again the role of the method practitioner is the one of a knowledge engineering, whose ability is to collect and map specific-domain knowledge to method's mechanisms.

Finally, analogies can be made with the best practice in risk management, in particular risk identification and analysis. During the phase of risk identification, the professional in charge of the process has to find evidence of potential threads. In order to accomplish this task some guidelines are available that suggests ways to identify threads. Guidelines could be a risk-based taxonomy, or basic technique such as observations, interview, past history of accidents, similar projects risk analysis. These generic techniques can be used to sustain that a domain element is a potential risk, exactly how trust scheme are generic technique to sustain that an element could be an evidence for trust.

The second stage, the risk assessment or quantification, is similar to the phase of trust computation where identified evidence is quantified.

It is not a surprise that risk assessment and trust assessment can share a similar practice, since they are closely related concept. Our method is therefore a tool to guide a trust analysis, in a similar manner as a risk analysis is performed. Due to its nature, close to the mentioned applications (*data-mining*, *expert-system*, *and risk analysis*) the system shares their benefits and weaknesses, as discussed shortly in the section dedicated to the problem of mapping.

7.3.3 Features of the computation

Application-Contained nature of the computation

The evaluation performed did not use any information coming from outside the application under analysis. In all the experiments (except II where nevertheless the system is defeated) we do not use rating or recommendations, used to build the alternative ranking to be compared with our computation.

This choice was justified for evaluation reasons, but it also adds to our computation an application-contained aspect. We computed trust depending entirely on internal elements of the application connected to its core functionality and not inserted for explicitly support trust computation, such as recommendations system or a dedicated trust infrastructure such as PKXI. Our computation results application-contained, with the benefit of being less invasive, in the sense that only available elements of the domain are used, not requiring a dedicated infrastructure to be added to application.

Justifications

An interesting feature of our method is the ability of giving justification. Justifications are given since the mechanisms used to take a decision are explicit in our trust scheme. Moreover, the argumentation layer can also provide reasons why an entity has been defeated. Justifications makes the decision-making process more transparent for humans, and, thanks mainly to the argumentation layer, gives a better understanding of specific cases and decisions taken.

Argumentation Scenario

We conclude underlying the original argumentative feature that our method may support. At the core of the method there could be a dialectical process between a trustier and a trustee, or a set of agents. As we discussed earlier, setting our method in a dialectical scenario makes it feasible in distributed scenarios, and makes each counterpart responsible for the accomplishment and the validity of the stages of the method, from data gathering to trust scheme mapping and plausibility study. In this scenario the value added by our method is the set of trust scheme and critical questions, representing the proper argumentative tools to sustain the conversation.

Alternative to the couple past-outcomes/Recommendation

Our computation exhibits a novelty by not relying on the protected territory of pastoutcomes paradigm or recommendation. This computation does not reduce trust to a probability, but it keeps a rich notion of human trust encoded in the trust scheme.

Our evaluation showed how the computation is efficient as reliable recommendation systems (the one used for the comparative evaluation), and it can be even more effective in some scenarios, as experiment II and III show.

In all this applications, our method could be an alternative to traditional ones, mainly based solely on recommendations or outcome-based computation, an alternative that can produce better results, confirm their results (used therefore as a check metric), or be the only method to be feasible, since environmental constrains prevent the collection of recommendations and outcomes.

Evidence Selection

We also claim that our computation exploits a larger set and novel evidence. In particular, time-based trust schemes, the use of personal information, similarity to a standard and stability. Again, these pieces of evidence can be used since there are assumed to be defeasible. We think that defeasibility is a key concept able to extend trust evidence, otherwise potentially harmful and invalid.

7.3.4 Open Issues and Future Development

The phase of Domain Analysis

This is the first stage of our method, where a complete model of the application has to be produced, showing all the elements that will be used for the matching and testing of trust schemes. In our evaluation we used a UML model of the application, but we wonder if, in order to enhance the process of domain analysis and matching of the trust scheme, we could define a modelling language for trust purposes, able to sustain a *trust-aware model* of an application. We envisage an augmentation of a basic representation of the domain. The representation procedure requires adding some relevant information to a basic domain representation. This procedure could be seen as a method to follow in order to describe a domain for trust purposes. Having defined the a priori trust schemes, this description is actually a gathering and a pre-matching of such schemes over domain elements, matching that will be quantified and tested in the next stage

using the critical questions. We may wonder why the procedure is needed and if it would not be enough to directly try to match our schemes on a domain representation. We identify four rationales: (i) setting up a general procedure requires less or no knowledge of each of the schemes; (ii) direct matching can be more prone to error and partial visions; (iii) a general procedure can be processed by chain-reasoning, whose conclusions could be hard to identify with a direct matching; (iv) a trust-aware domain representation is by itself a useful new contribution to trust studies. The goal of the modelling phase is to gather all the information needed to support the presumptive identification and the critical analysis of trust instances. The information includes:

- analysis of entities: properties, actions, recognisability, relationships, dependency, topology, goals
- analysis of actions: effort to complete, effect for the entities involved, dependency among actions, consequent actions, impact on entities, impact on possible outcomes of an action, observability of the outcomes
- how communication is possible (a requisite for making indirect trust more plausible)
- analysis of the environment: its dimension and how it changes in time (support persistence)
- memory constraints regarding entities, actions, objects

Mapping, Critical Question and Computation

One of the main issues of our method is the dependency of the results on some critical stages of the method, i.e. the mapping of trust scheme and critical questions, the computation performed over the selected evidence including the argumentation layer.

A similar discussion can be performed for the stage of defeaters analysis, where initial level of plausibility has to be decided in order to carry on the reasoning. This stage is critical to the correct application of the method and requires a certain amount of ability by the *trust-engineer*, the professional in charge of the method.

Which is the value added by the method, if the results are – maybe highly – dependant on mapping and computation?

It is doubtless that a good mapping leads to a good computation; that a practitioner could be more skilled than another in applying the process and that a bad mapping deteriorates the results. The problem is therefore investigating the process of mapping, identifying the tools our method provides to help the practitioners.

The discussion of this issue - we name *mapping* in the rest of the discussion - allows us to underline strengths and weaknesses of our methodology.

The problem of mapping as the problem of evidence selection in Trust

The first argument to use in the discussion is that the *mapping* problem is shared by all the trust solutions, and it is still an open problem. In this respect, our method offers improvements to the current solutions since it can be seen as a *best-effort* in the definition of an evidence selection strategy.

As discussed in the introductive chapter, the problem of evidence selection is solved in trust in the following ways:

- 1 by building a dedicated trust infrastructure where evidence are well-defined objects,
- 2 by solely relying on users judgments, that means that implicitly the whole process is delegated to users, and
- 3 by relying and again relegating the process to domain expertise.

We start by underline the difference between the first two solutions and the last one. In the first two solutions, the pieces of evidence used are well-defined objects defined *a priori* explicitly for support a trust computation. The problem of evidence selection and mapping is therefore well-bounded, while in the last example is clearly *unbounded* and potentially affected by subjectivity and lack of systematicity.

When evidence has to be collected over application elements, many trust solutions delegate this to the domain expert that identifies and produces the correct evidence for a specific domain. We note how the trust engine involved in the process does not offer any clues about how to collect evidence, and the reliance on domain expertise – its existence and availability – is strong. The value of trust as a distinct expertise is lost.

Other solutions perform trust computation using heuristics, with the problem of lack of systematicity and subjectivity. Even the lack of a trust-related meaning can be part of problem associated to heuristics.

Our solution: decomposition and domain dependency

Mapping is an unbounded and messy problem when trust has to be computed over domain elements. What can our model add to the solution of this issue?

Our model faces the problem of evidence selection as a mapping between trust schemes and domain elements. The process involves also answering to some defined questions that requires a further mapping and collection of domain elements.

Role of domain-specific knowledge in mapping

The first thing to underline is that our method contains actually a set of tools to perform evidence selection and plausibility study. Even if a good mapping could require an advanced understanding of the application and the participation of a domain expert, the difference is now that the solution is not delegated to the domain expert, but he is part of a process whose stages have been defined by a generic expertise of trust.

Each trust scheme defines situations that have to be matched over domain elements, and domain expert can help accomplishing this task, but they are not involved and responsible in the definition of a trust model for the specific domain (already contained in the trust schemes), but rather they guarantee its correct application. The individual responsible for indicating elements to be mapped with trust scheme is required to have an understanding of the application, not of trust or cognitive models, while the method practitioner is an expert of trust, not of the application.

Decomposition

Moreover, we noted how each trust scheme faces a limited and focused problem, which is not explicitly related to trust. For instance, the trust scheme activity focuses on quantitative aspects of an entity contribution. It does not require any assessment of the trustworthiness of such activity, but rather a classification of different types of activity, if required. The same stands for the trust schemes related to time: it is only the time distribution of the activity that is required, not qualitative assessments. We designed the trust schemes in order to isolate a specific issue, and it is responsibility of the argumentation layer to compose them.

Of course some schemes or critical questions require skill to be answered, but again they pose a specific problem. In this sense, schemes help to decompose the problem of trust into sub-problem that may be complex but not requiring knowledge about trust models.

We underline again how a good mapping is not between elements and trust, but between elements and what the trust scheme requires. Therefore, a good mapping of activity is the one that clearly identifies when an entity is more active than another. This led us to believe that a good mapping between trust scheme and elements is a reasonably feasible task.

Providing guidelines to facilitate the mapping

Moreover, guidelines about how to perform a matching can be defined. First of all, we notice that the schemes must be matched over the available domain elements. A good practice is to go through the elements of the applications, and wondering if they can be associated to some trust scheme, as we did in the *FinanzaOnline* experiment. The important point is that the problem to be matched should be clear and isolated, in order to avoid both a bad mapping and a mapping not relevant to the trust scheme.

Justification to keep a transparent mapping

Finally, as already stated above, trust schemes justify why an element is selected as trust evidence and makes the process transparent and re-tractable. This is different from delegating the process to a domain expert that may produce results that are not easy to explain, justified by some depth expertise, hard to make explicit and sometime personal and confidential. This means that the reason why an entity should be trusted could be not understandable, and it could be impossible to reject or agree with the underlying reasons.

Parallel with other best practice

The mapping-dependency is not only a problem of trust models. As we described above, our model can be seen as an expert-system for trust, or *trust analysis* similar to what it is done with risk analysis.

Therefore, the problem discussed of mapping and evidence selection has similarities with these other class of applications. Even in expert-systems the quality of the solution proposed relies on the ability of the knowledge engineer to model properly the required expertise. Moreover, during the application of the system a good solution is dependent on the quality of the input provided. If the input provided by the users is incomplete or they do not describe correctly the problem, the solution will be affected.

Good expert-systems limit this dependency by providing carefully designed input interfaces, with clear guidelines and examples. Moreover, domain expertise could be invoked in the process, or the practitioner could use any fact-finding techniques (interview, observation, sampling and questionnaire) exactly like a knowledge engineer or a risk analyst does as part of their analysis.

Learning approach

A learning approach could also be employed, that will marry the use of defeasible reasoning with a theory of beliefs revision such as the Bayesian approach, already well-known in trust.

One of the weak points of the present stage of the method – that does not claim to be exhaustive – is that initial plausibility values are highly subjective. Plausibility values and reasoning link's strength could be dynamically adapted when starting believes are confirmed or contradicted. While defeasible reasoning combines and organizes beliefs, a learning approach adjusts reasoning links strength. The agent learns how to reason defeasible and refine the importance given to each argument.

Uncertainty treatment

Another aspect to be considered is whether or not the method supports the treatment of uncertainty. A practitioner could be not sure about part of its mapping, or the data available uncertain. Therefore a treatment of the uncertainty should be embedded in the computation. Our method addresses this problem by treating uncertainty as a defeater whose strength in many cases can be statistically quantified.

Conclusions

In this final chapter we started by reformulating the objectives of our thesis, and by analyzing how the various issues where accomplished in our work, underlying strengths and weaknesses. We then analyzed the main contribution of our works, that can be summarized as the introduction

of defeasibility in trust computation and the definition of a novel computational approach to the classical recommendation and past-outcomes trust systems. In the final sessions we analyzed the featured of the method stressing future works and open issues.

Bibliography

Relevant Author's publications

[Don07b] P. Dondio, L. Longo. A translation Mechanism for Recommendations. *Proceedings of the 2nd IFIP joint conference on Trust Management*, Trondheim, Norway, June, 2008

[Sei07] JM. Seigneur, P. Dondio. Trust in self-organizing systems, chapter of the book *Self-organizing systems* edited by Springer. To be published December 2009

[Lon07] L. Longo, P. Dondio, S. Barrett, Temporal Factors to evaluate trustworthiness of virtual identities, proceedings of *IEEE SECOVAL 2007, Third International Workshop on the Value of Security through Collaboration, SECURECOM 2007*, Nice, France, September 2007

[Don07a] P. Dondio, E. Manzo, S. Barrett, Applied Computational Trust in Utilities Management a Case Study on The Town Council of Cava dei Tirreni, in *Trust Management, proceedings of IFIPTM, the first joint iTrust and PST Conferences on Privacy, Trust Management and Security*, Springer, July 2007.

[Don07c] P. Dondio, S. Barrett, Computational Trust in Web content quality *Informatica Journal*, N. 31, pages. 151-160, June 2007

[Don07d] P. Dondio, S. Barrett. Application-Contained Trust Calculation: a non Invasive Approach Based on Presumptive Reasoning and Intuitive Trust. *UbiSafe, IEEE Symposium on Ubisafe Computing*, Niagara Falls, 2007, Canada

[Don07e] P. Dondio, S. Barrett, Presumptive Selections of Trust Evidences. AAMAS 2007, 6th joint international conference on Multi-Agents and Autonomous Systems, Hawaii, 2007, USA

[Don06a] P. Dondio et al. Extracting Trust from Domain Analysis: a Study Case on the Wikipedia Project, *IEEE Automatic Trusted Computing Conference*, LNCS, 2006, Wuang, China

[Don06b] P. Dondio, S. Barrett, S. Weber: Calculating the Trustworthiness of a Wikipedia Article Using DANTE Methodology, *Proceedings of IADIS eSociety conference* 2006, Dublin

References and Bibliography

[Abd00] A. Abdul-Rahman, S. Hailes. Supporting trust in virtual communities. Proceedings of HICSS, 2000

[Abd97] A. Abdul-Rahman. A distributed trust model. *Proceedings of New Security Paradigms Workshop*, UK, 1997.

[Abe04] K. Aberer. Efficient, handling of Identity in a P2P network. *IEEE Transaction on Knowledge Engineering*, Vol.16, Issue 6, 2004

[Ale95] J. Alexander, M. Tate. Web Wisdom: How to Evaluate and Create Information Quality on the Web, Lawrence Eribaum Associates Inc, New Jersey, USA, 1995

[Ama02] Amazon Auctions, website: www.amazon.com Accessed on the 24th August 2007

[Ant07] A System for Modal and Deontic Defeasible Reasoning, AI 2007: Advances in Artificial Intelligence, Springer Berlin, pg. 609-613, Australia, 2007

[Ari28] Aristotle. The works of Aristotle Translated into English. Ed. W. Ross. Oxford University Press, United Kingdom

[Bal03]. E. Ball, D. Chadwick, A. Basden. The implementation of a system for evaluating trust in a pki environment. *Proceedings of Trust in the Network Economy. Evolaris*, 2003

[Bar05] K. Barber, J. Kim. Belief Revision Process based on Trust: Simulation engine. *Proceedings of iTrust05*, pp. 397–401, 2005

[Bar08] Barnstars Wikipedia award, http://en.wikipedia.org/wiki/Wikipedia:Barnstars. Last accessed on the 12th October 2008

[Bet94] T. Beth, M. Borchedring, B. Klein. Valuation of trust in open networks. *Proceedings of the European Symposium on Research in Computer Science (ESORICS)*, 1994.

[Bla96] M. Blaze, J. Feigenbaum, J. Lacy. Decentralized trust management. In *Proceedings of the IEEE Conference on Security and Privacy*, 1996.

[Bla96] M. Blaze, J. Feigenbaum, J. Lacy. Decentralized Trust Management. *Proceedings of IEEE Conference on Security and Privacy*. 1996

[Bre97] G. Brewka, J. Dix and K. Konolige. Non monotonic Reasoning, an overview. CSLI publications, Center for the Study of Languages and Information, Stanford University. United States, 1995

[Bry05] C. Bryce, P. Couderc, J.M. Seigneur, V. Cahill. Implementation of the secure trust

[Buc02] H. Bucher. Crisis Communication and the Internet: Risk and Trust in a Global Media. In *First Monday*, Volume 7, Number 4, 2002

[Bur00] R. Burke. Knowledge-based Recommender Systems. In *Encyclopaedia of Library and Information Systems*, vol. 69, Supplement 32, New York, 2000

- [Cah03] V. Cahill, et al. Using Trust for Secure Collaboration in Uncertain Environments. *IEEE Pervasive Computing Magazine*, vol. 2, N. 3, Special Issue July-September 2003
- [Car02] J. Carter, E. Bitting, A. Ghorbani. Reputation Formalization for an Information-Sharing Multi-Agent System. *Computational Intelligence* 18(2), pagg. 515-534. 2002
- [Car05] Carbone, Nielsen M., M. Sassone V. A Formal model of Trust in Dynamics Network. iTrust 2005, 3rd conference on Trust Management, Roquefort, France, 2005
- [Carb02] J. Carbo, J. Molina, J. Davila. Comparing predictions of SPORAS vs. a Fuzzy Reputation Agent System. Proceedings of the 3rd International Conference on Fuzzy Sets and Fuzzy Systems, pp. 147-153.Interlaken, Germany, 2002
- [Cas00] C. Castelfranchi, R. Falcone. Trust is much more than subjective probability: Mental components and sources of trust. *32nd Hawaii International Conference on System Sciences*. 2000.
- [Cas03] C. Castelfranchi, R. Falcone, G. Peluzzo: Trust in information sources as a source for trust: a fuzzy approach. *Proceedings of the first conference on Autonomous Agent and Multi-Agent System.* ACM Press, New York, NY 10036, USA, 2003
- [Cas95] R. Cassel. Selection Criteria for Internet Resources. College and Research Library News, N. 56, pagg. 92-93, 1995
- [Cas98] C. Castelfranchi, R. Falcone. Principles of Trust for MAS: Cognitive Anatomy, Social Importance, and Quantification. *Proceedings of the International Conference on Multi-Agent Systems (ICMAS'98)* pp.72-79. Paris, France. 1998
- [Chr96] B. Christianson, S. Harbison. Why isn't trust transitive? *Proceedings of the Security Protocols International Workshop*, University of Cambridge, pages 171–176, 1996
- [Chu96] Y. Chu. Trust Management for the World Wide Web, Master Thesis, MIT, 1996
- [Cio96] T. Ciolek. Today's WWW, Tomorrow's MMM: The specter of multi-media mediocrity, *IEEE COMPUTER Magazine*, vol. 29(1) pp. 106-108, January 1996
- [Cis07] R. Cissee, S. Albayak. An agent-based approach for privacy-preserving recommender systems. *Proceedings of the sixth conference on Autonomous Agent and Multi- Agent Systems*. ACM Press, New York, NY 10036, USA, 2007
- [Con02] R. Conte, M. Paolucci. Reputation in Artificial Societies: Social Beliefs for Social
- [Del03] C. Dellarocas. The digitalization of Word-Of-Mouth: Promise and Challenges of Online Reputation Mechanisms. *Proceedings of Business Process Management*, 2000
- [Des04] Z. Despotovic, K. Aberer. Maximum likelihood estimation of peers performance in p2p networks. *Proceedings of the Second Workshop on the Economics of Peer-to-Peer Systems*, 2004

[Deu49] M. Deutsch. A theory of Cooperation and Competition. *Human Relations*, 2(2), 129–152, 1949

[Deu62] M. Deutsch. Cooperation and Trust: Some Theoretical Notes. *In: Jones, M. R. (ed), Nebraska Symposium on Motivation*. Nebraska University Press, 1962

[Eba02] eBay website: http://www.eBay.com

[Eco06] e-Consultancy Web site. Poor accessibility can affect customer trust. Retrieved at http://www.e-consultancy.com/news-blog/360883/poor-website-accessibility-can-affect-customer-trust.html on the 12th February 2007 Last Update 12 January 2006.

[Epi00] Epinios, www.epinions.com. Accessed on the 12th May 2007

[Esf01] B. Esfandiari, S. Chandrasekharan. On How Agents Make friends: Mechanisms for Trust Acquisition. *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada. pp. 27-41, 2001

[Fal04] R. Falcone, C. Castelfranchi: Trust Dynamics: How Trust Is Influenced by Direct Experiences and by Trust Itself. *AAMAS 2004:* pp. 740-747, 2004

[Fen04] Fenton, N., Neil, M.: Combining evidence in risk analysis using Bayesian networks. Tech. rep., Agena (2004)

[Fie00] R. Fielding. Architectural styles and the design of network-based software architectures. PhD dissertation, University of California, Irvine, 2000 [Fin08] FinanzaOnline, finance and trading portal. Web site: www.finanzaonline.it Last accessed on the 12th August 2008.

[FOA07] FOAF project website: www.foaf-project.com. Accessed on the 16th of August 2007

[Fog03] B. Fogg. How Do Users Evaluate The Credibility of Web Sites? *Proceedings of the conference on Designing for user experiences*, ACM Press, USA, 2003

[Fre03] L. Frewer, S. Miles. Temporal stability of the psychological determinants of trust: Implications for communication about food risks. Health, Risk and Society 5, 259–271(13), 2003

[Fri99] E. J. Friedman, P. Resnick. The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy* 10(2): 173-199, 1999.

[Fud89] D. Fudenberg, D. Levine. Reputation and equilibrium selection in games with a patient player. *Econometrica* 57(4), 1989

[Gab02] S. Gabbay, R. Leenders. A perceptional view of the Coleman model of trust. 2002. Retrieved online at http://www.ub.rug.nl/eldoc/som/b/02B32/02B32.pdf, Accessed on the 24th May 2006

- [Gal05] J. Gales. Encyclopaedias goes head a head, Nature Magazine, issue N. 438, 15, 2005
- [Gam00] D. Gambetta. Can we trust trust? Chapter from the book *Trust: Making and Breaking Cooperative Relations*, pp. 213–237, 2000. Published at www.sociology.ox.ac.uk/papers/gambetta.pdf. Accessed on the 17th April 2006
- [Gar06] A. Garg, A. Montresor, R. Battisti: Reputation Lending for Virtual Communities. *ICDE Workshops*. 2006
- [Gil05] J. Giles. Special report: Internet encyclopaedias go head to head. *Nature* (438), 900–v901, 2005
- [Gol02] J. Golbeck, J. Hendler, B. Parsia. Trust Networks on the Semantic Web, University of Maryland, USA, 2002
- [Gol04] J. Golbeck, J., Parsia, B.: Trusting claims from trusted sources: Trust network based filtering of aggregated claims. *Proceedings of the 3rd International Semantic Web Conference*, LNCS 3298. Springer-Verlag, 2004.
- [Gon06] J.M. Gonzalez-Barahona, M. Conklin, G. Robles. Public data about software development. *Proceedings of the International Conference on Open Source Software*, 2006
- [Gra00] T. Grandison, M. Sloman. A survey of trust in Internet application. *IEEE, Communications Surveys,* Fourth Quarter, 2000.
- [Gra06] L. Gray. Trust-Based Recommendation Systems. PHD Thesis, Trinity College Dublin, Ireland, 2006
- [Guh04] R. Guha. Open rating systems. Technical report, Stanford University, USA, 2004
- [Hua04] J. Huang, S. Fox. Uncertainty in knowledge provenance. *Proceedings of the first European Semantic Web Symposium*, Heraklio, Greece, 2004.
- [Jos05] A. Josang, S. Pope. Semantic Constraints for Trust Transitivity. *Proceedings of the SecondAsia-Pacific Conference on Conceptual Modelling*, Newcastle, Australia, 2005
- [Jos96] A. Josang, The right type of trust for distributed systems. *Proceedings of the New Security Paradigms Workshop*. ACM (1996). URL http://citeseer.nj.nec.com/47043.html
- [Jøs99] A. Jøsang. An algebra for assessing trust in certification chains. In *Network and Distributed Systems Security*, 1999.
- [Jose02] S. Joseph. Neurogrid: Semantically routing queries in peer-to-peer networks. In *International Workshop on Peer-to-Peer Computing*, Pisa, Italy, May 2002.

[Kau03] R. Kauffman, C. Wood. Detecting, predicting and preventing reserve price shilling in online auctions. *International Conference on E-Commerce*, Pittsburgh, USA, 2003.

[Koh08] Kohlas R., Jonczy J., Haenni R. A trust Evaluation Method Based on Logic and Probability Theory. Trust Management II, proceedings of IFIPTM 2008, Trondheim, Norway, 2008

[Kre82] D. Kreps, R. Wilson. Reputation and imperfect information. Journal of Economic Theory 27, 1982

[Lag96] O. Lagerspetz, O. The Tacit Demand. A Study in Trust. Filosofiske Institutionen. Abo, 1996

[Lew95] J. Lewicki, B. Bunker. Trust in relationships: A model of development and decline. In J. Z. Rubin (Ed.) Conflict, cooperation, and justice: Essays inspired by the work of Morton Deutsch. San Francisco, 1995

[Luh00] N. Luhmann. Familiarity, confidence, trust: Problems and alternatives. Chapter from the book *Trust: Making and Breaking Cooperative Relations*, pp. 213–237, 2000. Published at www.sociology.ox.ac.uk/papers/gambetta.pdf. Accessed on the 17th April 2006

[Mar05] S. Marsh, Trust, Untrust, Distrust and Mistrust - An Exploration of the Dark(er) Side. *Proceedings of iTrust 2005, 3rd conference on Trust Management*, Roquefort, France, 2005

[Mar94] S. Marsh. Formalizing Trust as a Computational Concept. PhD thesis, University of Stirling, Scotland, 1994

[McA95] D. J. McAllister. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38, 24-59, 1995

[Mas05] P. Massa, P. Avesani. Controversial users demand local trust metrics: An experimental study on epinions.com community. *Journal of AAAI*, pp. 121–126, 2005

[McG06] D. McGuiness, H. Zeng, P. da Silva, L. Ding, D. Narayanan. Investigations into trust for collaborative information repositories: A Wikipedia case study. *Proceedings of the WWW2006 Workshop on the Models of Trust for the Web (MTW'06)*. ACM Press, New York, NY 10036, USA, 2006

[McK00] D. McKnight, N. Chervany. What is trust? a conceptual analysis and an interdisciplinary model. *Proceedings of the Americas Conference on Information Systems*, 2000

[McK02] D. McKnight, L. Chervany. Notions of Reputation in Multi-Agent Systems: A Review. *Proceedings of the 34th Hawaii International Conference on System Sciences*. Honolulu, Hawaii, USA. 2002

[McK96] D. McKnight, L. Chervany. The meanings of trust. Technical report, University of Minnesota Management Information Systems Research, 1996

[McQ67] J. MacQueen. Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics, Berkeley,* University of California Press, USA, 1967

[Mes07] A. Messery. Expectations enhanced trust value. *Proceedings of the Ninth Workshop on Trust in Agent Societies*, pp. 70–77, Hawaii, USA, 2007

[Ons02] Onsale Exchange. Website www.onsale.com. Accessed on the 17th July 2006

[Ope03] Sierra, OpenPrivacy: OpenPrivacy reputation management framework. http://sierra.openprivacy.org. Accessed on the 15th May 2005

[Pag99] L. Page, S. Brin. The PageRank citation Ranking: bring Order to the Web, Technical Report, Standford University, US, 1999

[Pea88] J. Pearl. 1988, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, 1988

[Pic76] T. Pickett, L. Sussman. Causal Attributions and Perceived Source Credibility: Theory, Data, and Implications, *ERCIM Database*, *www.ercim.org*. *Code ED131509*, 1976

[Pie58] Collected Papers of Charles Sanders Pierce, Harvard University Press, Hartshorne et al.

[Pin07] I. Pinyol, S. Jordi, C. Guifre. How to talk about reputation using a common ontology: From definition to implementation. *Proceedings of the Ninth Workshop on Trust in Agent Societies* pp. 90–102, Hawaii, USA, 2007

[Pol01] J. Pollock. Defeasible Reasoning with Variable Degrees of Justification. *Artificial Intelligence*, N. 133, pages 233-282, 2001

[Pol94] J. Pollock, Justification and defeat, Artificial Intelligence, N. 67, pages 377–408, 1994

[Que06] D. Quercia, M. Lad, S. Hailes, L. Capra, S. Bhatti. STRUDEL: supporting trust in the dynamic establishment of peering coalitions. Proceedings of ACM SAC06, pp. 1870-1874, 2006

[Rah05] A. F. Rahman. A framework for decentralized trust reasoning. PhD dissertation, University of London, 2005

[Rei80] R. Reiter, A logic for default reasoning, *Artificial Intelligence*, N. 13, pages 81–132, 1980

[Ric03] J. Rice. Mathematical Statistics and Data Analysis. Duxbury Press, 2nd Edition. Belmont, California, USA, 2003. ISBN 0534209343

[Rie08] S. Ries, A. Heinemann. Analyzing the Robustness of CertainTrust. In Trust Management, LNCS, pagg 51-67, Springler, ISBN 978-0-387-09427-4, 2008

[Rob04] T. Roberts. Online Collaborative Learning, Theory and Practice, Idea Group Pub, USA, 2004

[Rom03] D. Romano. The Nature of Trust: Conceptual and Operational Clarification, Louisiana State University. PhD Thesis, 2003

[Sab01] J. Sabater, C. Sierra. REGRET: A reputation model for gregarious societies. *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada. pp. 61-69, 2001

[Sab02] J. Sabater, J. C. Sierra. Reputation and Social Network Analysis in MultiAgent Systems.. *Proceedings of the first international joint conference on autonomous agents and multiagent systems* (AAMAS-02), pp. 475-482. Bologna, Italy,

[Sab05] J. Sabater, C. Sierra. Review on computational trust and reputation models. In *Artificial Intelligence Review*. Vol. 24, Num. 1, pages 33-60, September 2005.

[Sab06] J. Sabater. Towards Next Generation of Computational Trust and Reputation Models. In Modelling Decisions for Artificial Intelligence, LNCS, vol. 3885, pages 19-21, February 2006.

[Sch00] M. Schillo, P. Funk, M. Rovatsos. Using Trust for Detecting Deceitful Agents in Artificial Societies. *Journal of Applied Artificial Intelligence (Special Issue on Trust, Deception and Fraud in Agent Societies)*, 2000

[Sei04] J.M. Seigneur, C. D. Jensen. Trading privacy for trust. *Proceedings of Trust 2004*, pp. 93–107, 2004

[Sei05] J.M. Seigneur. Trust, security and privacy in global computing. PhD dissertation, Trinity College Dublin, Ireland, 2005

[Sei06] J.M. Seigneur. Ambitrust? Immutable and Context Aware Trust Fusion. Technical Report, University of Geneva, 2006

[Sen02] S. Sen, N. Sajja. Robustness of Reputation-based Trust: Boolean Case. *Proceedings of the first international joint conference on autonomous agents and multiagent systems* (AAMAS-02), pp. 288-293. Bologna, Italy, 2002

[Sha76] Shafer, Glenn. A Mathematical Theory of Evidence, Princeton University Press, 1976, ISBN 0-608-02508-9

[Spo02] W. Spohn. A brief Comparison of pollock's Defeasible Reasoning and Ranking Functions. In *Synthese Journal*, Volume 131, Num. 1, pages 39-56, April 2002

[Sta05] Standford Encyclopedia of Philosophy. Defeasible Reasoning entry retrieved at http://plato.stanford.edu/entries/reasoning-defeasible. Accessed on the 27th August 2008.

[Sta06] Standford Web Credibility Guidelines. Retrieved at http://credibility.stanford.edu/guidelines. Accessed on March 2006

[Str08] Stranders R., Weerdt M., Witteveen C.Fuzzy Argumentation for Trust. Computational Logic in Multi-Agent Systems: 8th International Workshop, pages 214-230, published by Springer-Verlag, Berlin, 2008 isbn 978-3-540-88832-1

[Szt00] P. Sztompka. Trust: a Sociological Theory. Cambridge University Press. Cambridge, UK, 2000

[Ter04] S. Terzis, W. Wagealla. The SECURE Collaboration Model, Technical Report, Trinity College Dublin. Deliverables 2.1, 2.2 and 2.3 available at http://secure.dsg.cs.tcd.ie. Accessed on May 2007

[Tou84] D. Touretzky, Implicit orderings of defaults in inheritance systems, in *Proceedings of AAAI-84*, Austin, TX, 1984.

[Tru04] Trustcomp, Computational Trust Online Community. http://www.trustcomp.org. Accessed on the 23 February 2006

[Tve74] A. Tversky, D. Kahneman. Judgment under uncertainty: Heuristics and biases. In *Science*, New Series 185(4157), 1124–1131, 1974

[Tve82] P. Tversky, D. Slovic. Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, UK,1982

[Wal96] D. Walton. Argumentation Schemes for Presumptive Reasoning.. Lawrence Erlbaum Associates, United States, 1996 - ISBN 080582071X, 9780805820713

[Wan03] Y. Wang, J. Vassileva. Bayesian network trust model in peer-to-peer networks. *Proceedings of AP2PC03*, pp. 23–34, 2003

[Wik06a] Online article *How much do you trust Wikipedia?* Retrieved from http://news.com.com/20091025_3-5984535.html, accessed on 5th March 2006.

[Wik06b] Wikimedia, repository site of the Wikipedia project. http://download.wikimedia.org, accessed on the 17th March 2006

[Wik06c] Wikipedia Online Encyclopedia. URL: http://www.wikipedia.org. Accessed on the 15th July 2008

[Wua06] Y. Wuang. Bayesian Network Based Trust Management, *Proceedings of IEEE ATC06*, Wuhan, China, 2006

[Yu01] B. Yu, P. Singh. Towards a Probabilistic Model of Distributed Reputation Management. *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, Montreal, Canada. pp. 125—137, 2001

[Yu02] B. Yu, P. Singh. An Evidential Model of Distributed Reputation Management. *Proceedings of the first international joint conference on autonomous agents and multiagent systems* (AAMAS-02), pp. 294-301. Bologna, Italy, 2002.

[Zac99] G. Zacharia. Collaborative Reputation Mechanisms for Online Communities. Master's thesis, Massachusetts Institute of Technology, 1999

[Zen06] H. Zeng. Computing Trust from Revision History, *Proceedings of PST 2006, international conference on Privacy, Security and Trust, Canada, 2006*

[Zie05] C. Ziegler, J. Golbeck. Investigating Correlations of Trust and Interest Similarity, *Decision Support Systems* 43(2), 460–475 (2007)

[Zim00] P. Zimmerman, PGP: Pretty Good Privacy, documentation retrieved from www.pgp.com

Appendixes

Appendix A Computational Models of Trust

This appendix contains a detailed description of the actual landscape of computational trust models. The appendix complements the discussion of chapter 2, that we kept at an higher level, focusing on the concepts behind each computational incarnation of trust rather than discussing the actual computational implementations, examples and techniques used, that is described in this appendix..

Computing trust using Past Outcome and Direct Experience

This computational mechanism uses evidence that the trustier gathered directly from previous interactions to predict trustee's future behaviours.

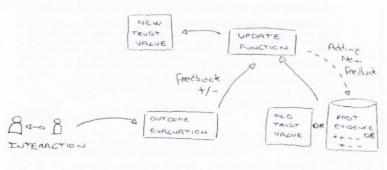


Figure A.1 Generic shape of the update loop of Direct Experience

In order to use the past outcome method, as shown in figure A.1, a trust model has to define an *update function* U, that computes a trust value post-interaction as a function of the trust value before the interaction and the outcome of the interaction just completed. This implies the presence of another function able to evaluate how good or bad was the interaction, the *feedback function* F. The function assignees to each of the possible outcomes a value representing the trustier level of fulfillment gained from that particular outcome. Therefore:

$$F: O \rightarrow Level of Satisfaction$$
 (A.1)

The feedback functions is defined from the set O of possible outcomes of an interaction to a value representing trustier level of satisfaction. F could be a collection of human explicit feedback or a result of an autonomic computation. The function of update is in general of this shape:

$$T_{vnew} = U(f, T_{vold})$$
 or $T_{vnew} = U(f, E_{old})$ (A.2)

U is a function that, given the old trust value and the level of satisfaction of the last interaction f

(the feedback), produces a new trust value. The old trust value T_{vold} could be replaced by the set of evidence E_{old} collected before the interaction, that joint with the new evidence f produces the new T_v .

As an example, we analyze the updating mechanisms of the CertainTrust model [Rie08]. We remind how a trust value is represented in this model as a couple (t,c) representing the level of trust t that A holds in B and the certainty level c associated with t. An opinion about an agent A is represented by a particular probability distribution b(r,s) defined between -1 (untrust) and 1 (trust), probability distribution that is a function of two parameters r and s, representing the number of positive (r) and negative (s) past interactions with an agent. The feedback function F has three possible values: -1 (negative), 0 (neutral) and 1 $(positive\ feedback)$.

After an interaction, the opinion (t,c) is updated by incrementing by one the value of r (if f=1), or s (if f=-1) or leaving it unchanged (f=0), with the consequence of changing the shape of the probability distribution b(r,s) representing the trust value. The shape of b(r,s) is a beta distribution of parameters r-1 and s-1, described in the next section.

In Hailes model [Abd00], each agent keeps count of three set of past evidence: very positive, neutral and very negative, whose update influences the agent's trust value trough a linear combination. Other model, like the one proposed by Quercia et Al. [Que06], update the new trust value using Bayesian theorem. Trust value has an associated probability function and, using as a posteriori condition the level of satisfaction of the interaction just concluded; the a priori trust value is update.

An important class of updating function is represented by linear combinations of past trust values and last interaction outcome represented by this general shape:

$$T_{v \, ngw} = m * T_{v \, old} + (1 - m) * f$$
 (15)

Where m represents the memory factor, i.e. how the last interaction affect the new trust value. In the extreme case, when m=1 the trust value is not dependant on past interactions, while when m=0 the trust value is totally dependent on the last feedback f and the agent does not keep any memory of the past.

A value between 0 and 1 represents a system that reacts to new interactions slowly (m small) or fast (m high). Usually, the choice of m depends on factors related to the environment: for high volatile environment where agents are likely to change rapidly, m should be chosen small, while in a stable environment m should be kept high to take advantage of past experience.

An example of such linear feedback system is present in the p2p-based trust model of Wang [Wan03], where the new trust value is equal to:

$$T_{v(i,j)_n} = a * T_{v(i,j)_0} + (1-a) * e_a$$
 (A.3)

Where e_a is the outcome of the interactions in [-1,1] and a is the learning rate, identical to the memory factor m discussed above.

In Abdul-Rahman [Abd00], each agent has a 4-tuple associated, containing the number of times an agent has been considered *very untrustworthy, untrustworthy, trustworthy, and very trustworthy*. The overall level of trust derived from past experience is computed by considering the maximum value of the tuple. For example, a tuple (0,3,2,4) will result in a *very untrustworthy* agent. When the maximum value is more than one, a level of uncertainty is attached to the trust value according to a table of possible situations.

Schillo [Sch00] computes the level of trust of an agent simply by updating the formula

e/n, where n is the total number of interactions and e is the number of positive interactions. The formula has a straightforward probability meaning.

Afras [Carb02] proposes an upgrade of the trust value of an agent using a weighted aggregation. The weights of this aggregation are calculated from a single value they call remembrance that behaves like a dynamic memory factor. This factor allows an agent to give more importance to the latest interaction or to the old reputation value. The remembrance factor is modeled as a function of the similarity between (1) the previous reputation of the trustee and the satisfaction of the last interaction and (2) the previous remembrance value. If the satisfaction of the last interaction and the reputation assigned to the partner are similar, the relevance of past experiences is increased by increasing the remembrance factor. If the satisfaction of the last interaction and the reputation value are different, the relevance of the last experience is increased by reducing the remembrance factor.

The direct experience evidence is without any doubt the most used computational trust mechanism. Anyway, there is too much confidence in its applicability. In Sabater's [Sab05] review of trust model, direct experience is regarded as *the most reliable source of trust*. We agree with this statement, since by relying on direct experience an agent should be sure of its level of satisfaction, reducing the noise due to the subjective differences that may arise by using others trust value. But the direct experience is, as any mechanism, a presumption that should be tested and that could be not so easy to implement.

First, the feedback function F could be not feasible and the outcomes not measurable and quantifiable. This implies that F is implemented in a too approximate way that can lead to incomplete or meaningless trust values. We note how usually many of these problems are bypassed relying on human explicit feedback. Other factors that should be considered before relying on this computational paradigm are:

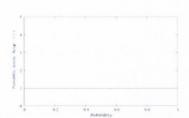
- the number of outcomes, their frequency and their temporal distribution that, may invalidates the trust value that becomes also subject to high variance,
- the stability of the situation/agents that may change, invalidating past evidence
- what it is possible to known about interactions' outcomes, which could be not observable or partially observable, making impossible or highly uncertain their evaluation and the consequent trust value.
- External constraints out of the control of the trustee that affects the outcome of an interaction, that should not affect trustee's reputation

The Theory of Probability as a Computational Tool in Trust

Probability techniques: beta function and Bayesian Update

Among the probability techniques used for computing trust, the *Bayesian* approach is the most popular. It is often coupled with the *beta*-distribution family of *pdf*, making the Bayesian update of beta distributions a classical example of the use of probability-based trust computation. The beta-distribution is a family of *pdf* used to represent a distribution over binary outcomes. A beta-distribution is completely defined by two positive numbers.

The two parameters define completely the expected value and the shape of the distribution. As an example, in fig A.2 is presented a beta distribution with the value of (1,1) on the left and (8,2) on the right.



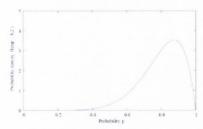


Figure A.2 Beta Distribution family.

This behavior maps directly with a representation of a trust value based on evidence. Usually, the two parameters a and b are the number of positive and negative evidence regarding the trustee, and the pdf distribution characteristics (expected value, variance) are used for trust values computation and uncertainty assessment.

The method is used by Josang [Jos05] or in the *CertainTrust* model [Rei08], where r and s are respectively the good and bad evidence regarding a trustee, and a=r+1 and b=s+1 define the corresponding beta pdf.

Referring to figure A.2, when no information is available about an agent, (r=s=0), the beta distribution (1,1) is uniform: no value is more likely than others and the uncertainty is at its maximum value. When, for instance, an agent holds 7 positive pieces of evidence and 1 negative, the corresponding beta distribution (8,2) is distributed around the average value of 0.8 with a small variance.

Beta distribution is an example of the advantages of the probability approach: clear meaning, well-defined and meaningful computations. As a tool for representing and processing trust, especially when coupled with direct or indirect feedback-based experience, it represents a well studied and meaningful model. However, manipulation of probabilities can, as pointed out by Castelfranchi and Falcone, being a reductive approach that can easily lose meaning by hiding the reasons for a trust decisions. A single probability value can contain, or hide arguments, reasons and evidence involved in the trust evaluation, that in many case are necessary for taking a good trust decision. Therefore, it should be coupled with other information, that, when needed, can be disclosed in order to support a better decision making process.

The *Bayesian update* is another popular mechanism used to update trust values. Bayesian analysis consists of formulating hypotheses on some real-world phenomenon, running experiments to test such hypothesis, and thereafter updating the hypotheses –if necessary– to provide a better explanation of the experimental observations, a better fit of the hypotheses to the observed behaviors. In trust, the hypothesis to be tested and eventually update is the trustee level of trustworthiness in the light of new evidence.

More precisely, Bayes' theorem adjusts probabilities given new evidence in the following way:

$$P(H_0|E) = \frac{P(E|H_0) \ P(H_0)}{P(E)}$$
 (A.4)

where

- H_0 represents a hypothesis, called a null hypothesis that was inferred <u>before</u> new evidence E become available.
- $P(H_{\theta})$ is called the prior probability of H_{θ} .
- P(E | H_0) is called the conditional probability of seeing the evidence E given that the hypothesis H_0 is true. It is also called the likelihood function when it is expressed as a

function of E given H_0 .

- P(E) is called the marginal probability of E: the probability of witnessing the new evidence E under all mutually exclusive hypotheses. It can be calculated as the sum of the product of all probabilities of mutually exclusive hypothesis and corresponding conditional probabilities.
- $P(H_0 \mid E)$ is called the posterior probability of H_0 given E.

An example in the context of trust can be the following. Suppose that the trustier has to interact with a trustee T. The trustier thinks that the trustee has a level of trust equal to H_0 =vt (very trustworthy) with probability $p(H_0)$, the prior probability based on all the knowledge gathered before the interaction. After the interaction, the trustier quantifies its level of satisfaction, and for instance he discovers that the trustee acted like an ut (untrustworthy) entity. Thus, the new evidence E is ut.

The trustier applies the Bayesian inference in order to update its beliefs in the trustee T.

In order to do this, he consider $P(E|H_0)$, the probability for an agent that is very trustworthy (H_0) to act like an untrustworthy one, and P(E), the probability of having in general an untrustworthy behavior from an agent. By applying the Bayesian rule, the trustier updates $P(H_0|E)$, the a posteriori (amended) probability of still having a very trustworthy agent after a very untrustworthy interaction E. In our example, the new amended probability is likely to result little, since the probability to have such an outcome if the prior hypothesis was true is little, since it is hard to sustain that a very trustworthy agent acts in a very untrustworthy way. The starting hypothesis of the agent being very trustworthy $P(H_0)$, is therefore adjusted after E.

Computing trust using Indirect Experience

Trust Network Representation

Common to any computational mechanisms based on *indirect experience* is the notion of a *trust graph*, where each node represents an agent and edges represent trust relationships from one agent to another (not symmetrical). A weight on the edge represents the trust value of the target agent according to the trustier's trust metric.

The graph should accommodate all the information needed for processing recommendations: timestamp, contextual information, numeric representations etc...

In the early work of Golbeck [Gol02], a pair of node is connected by a single oriented edge with a vale from 1 to 10.

In Histos [Zac99], each edge represents the most recent reputation rating given by one agent to another. The *root node* represents the agent owner of the graph.

In Schillo' *TrustNet* [Sch00] there is a directed graph where nodes represent witnesses and edges carry information about the observations that the parent node agent told about the child node agents to the owner of the net (the root node). Note how in Schillo's model recommended values are hierarchical since a parent node can report about its child node but not vice versa. In these two models, the graph topology changes in relation to the different trustier point of view.

Josang in [Jos05] performs a fine study over the transitivity of trust and its semantic constraints. Its trust network has two types of edges. The first one represents functional trust, the second referral trust. Functional trust is defined as the level of trust that an agents hold in another regarding a specific purpose and task, while referral trust is the level of trust in the trustee's ability to recommend a third agent for a specific task. Each trust relationship, functional or referential, is an edge labeled with a 4-tuple (type,m,p,t), where type is the type of relationship (functional or referral), m is the trust measure (the trust value in the selected representation), p is

the *trust purpose* (describing the situation) and t is the *timestamp* (of the last time the trust relationship was computed).

In some authors edges are not in relation with trust values but rather with interactions. In the work by Despotovic [Des04] there is a direct edge from A to B for each interaction occurred between the two agents, and each edge is marked with a couple (a,b) where a is either d (direct experience) or r (B gave a recommended value to A) and b is a real number in [0,1] expressing the outcome of the interaction.

In general, the trust graph follows the trust representation used by the author and contains all the set of information needed by the model to compute trust.

Aggregating: Sum, averaging and beliefs

The simplest form of aggregating function is the sum. EBay [Eba02] represents the most known and in many ways successful online reputation model. Its success has been proved by Dellacrosas' study [Del03] that concludes how this rating system has achieved its goal to increase the number of interactions among users, and increase their sense of confidence even in presence of strangers. In eBay, a single trust value (the *feedback*) is represented as a Boolean value of -1 or 1 or 0 (neutral) and they are all sum up to obtain a global value.

Amazon auctions [Ama02] and OnSale Exchange [Ons02] use a mean of all ratings to aggregate the trust values.

In general, the averaging approach tends to produce compact values that can synthesize a large number of information. The price to pay is a value that merges and blends conflicts and loses information potentially of critical importance. The average approach loses less information than the global sum, since the number of feedback is used in the computation. Thus, if we have the following extreme two cases:

- 1. 1002 feedback, 502 positive, 500 negative
- 2. 4 feedback, 3 positive and 1 negative

the EBay value is +2 in both the cases while the average is 0.02 and 0.5

In this case the average approach is more meaningful: it is more likely that in the second case the entity should be considered more reliable than the first one. The above is therefore an example of two trust computations, one more plausible than the other. In a defeasible argumentation, an agent has an argument to defeat the results of an EBay-like computation if extra counter argument (such as the number of feedback) are not provided by the other agent.

This is why the recent (March 2007) development of the eBay feedback system has added to the global seller trust value several additional information such as the number of feedback used to compute the rating and their temporal sequence.

In the model proposed by Abdul-Rahman and Hailes [Abd00], the information is combined by taking in consideration another factor, the similarity between the trustier agent and the recommender agent. In the specific model, the similarity refers to the way they judge agents they both know. The mechanism wants to address the uncertainty resulting in aggregating trust values from different agents that - even if all honest - have opposite point of view. Similar agents' opinions are given a higher importance while dissimilar agents' opinions are reduced or discarded. The idea behind this approach resembles *collaborative filtering*.

In [Yus02] the authors propose to combine different opinions using Dempster-Shafer theory of evidence, specifically the rule of combination. Dempster-Shafer [Sha76] is a mathematical theory of evidence based on *belief functions* which is used to combine separate independent pieces of information (evidence *E*) to calculate the probability (or *belief*) of an event *A*. Dempster-Shafer allows computing a confidence interval for an event *A*, deducted from

subjective probability of some other evidence E_i , all independent, but correlated to A. Given ϵ set of events X, a mass (or belief function) m over the set X is a function that verifies the following two properties:

$$m(\emptyset) = 0$$
 (A.5)
 $\sum_{Y \in P(X)} m(Y) = \mathbf{1}$ (A.6)

where P(X) is the power set of X.

In the Dempster-Shafer theory, given an event A, the probability P(A) is included in a confidence interval whose extremes are defined by the belief in A, represented by Bel(A), and its plausibility Pla(A). Therefore

$$bel(A) < P(a) < Pla(a)$$
 (A.7)

The belief brings together all the evidence that would lead us to believe in P with certainty. The plausibility brings together the evidence that is compatible with P and is not inconsistent with it. Formally, belief is the sum of the mass of all the evidence that are subset of A, while plausibility is the sum of the masses of all the evidence that intersects A.

Figure A.3 can help visualize the concept. We are interested in a confidence level for proposition A. Evidence $E_{I,2,3}$ are subset of A and they define the belief in the proposition A, it is certain that each E_i implies A. E_4 and E_5 has an intersection with A, and they are used to increase the plausibility of A, adding on the contrary more uncertainty, since their presence do not necessarily guarantee A.

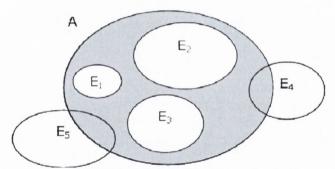


Figure A.3 Plausibility and Belief in A based on evidence E_i

The rule of combination is used when we need to combine two or more events. The combined mass m_{12} of two masses m_1 and m_2 of the events B and C is:

$$m_{12}(\emptyset) = 0$$

$$m_{12}(A) = \frac{\sum_{B \cap C = A \neq \emptyset} m_1(B) m_2(C)}{1 - K}$$

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$
(A.8)

We briefly comment the formula. K is a measure of the amount of conflict between the two mass sets, since it considers evidences not in common. If the two masses have nothing in common (intersection between B and C is the empty set), K = 1 and K-1 is zero and therefore combination of two disjoint events is impossible.

Therefore the combination rule ignores conflicts and it attributes any mass associated with

conflict to the null set. Conflicts are only a normalization factor that affects the value of the agreement between two events, on which the combination value depends.

If B and C are very similar K is large. This rule strongly emphasizes the agreement between multiple sources and ignores all the conflicting evidence through a normalization factor. Consequently, this operation yields counterintuitive results in the face of significant conflict in certain contexts and the use of that rule has come under serious criticism when significant conflict in the information is encountered.

In [Jos05] Josang performs a comparison between his subjective logic and Dempster-Shafer rule, underling how the counter-intuitive approaches of the last are corrected by his subjective logic.

Finally, Afras [Carb02] proposes a model where reputations collected are represented by fuzzy sets. Reputations are therefore matched using fuzzy-set overlapping, and the shape of the resulting fuzzy set represents the uncertainty associated with the aggregated reputation value. A narrow set means a reliable reputation value while a large set the opposite.

The problem of malicious ratings and recommender's trustworthiness

A significant class of reputation systems – such as eBay - accepts any feedback inserted without checking for malicious values. They rely on the fact that the number of ratings is usually high (they fit well big online communities) and therefore malicious or erroneous information will be diluted and have a smaller impact.

In general, reputation systems have some mechanisms in place to cope with the problem of malicious ratings and outliers.

The distribution of the ratings according to their values is a first important information. The shape of the distribution and its variance are often used to compute uncertainty of aggregated ratings or to spot suspicious values.

In the model by Afras [Carb02], the opinions from agents with bad reputation are not taken into account.

Sporas [Zac99] defines a measure of the reliability of the reputation based on the standard deviation of the ratings used to compute reputation. Standard deviation adds extra information about how the ratings are spread, and it could spot conflicts as increases in the variance value.

Histos [Zac99] introduces the concept of weighted recommended value according to the trustworthiness of the source. In his model, he uses the reputation of an individual as a measure of the reliability of its recommendations. Anyway, he does not make any difference between the ability of accomplishing a task and the ability to recommend a trust value. On the contrary, as we have mentioned above, Josang introduces the term *functional* trust and *referral* trust for these two concepts that are treated completely separated. In fact, a good recommender does not give any guarantee on its trustworthiness for other purposes.

The trust model by Sen and Sajja [Sen02] proposes a mechanism to assess the reliability of the recommendations that starts from evaluating how trustworthy is the environment as a whole, i.e. by estimating the percentage of liars in the systems. Knowing this fraction, the model computes how many agents should be queried to be sure that the likelihood of selecting a good partner has at least a specified value.

How to predict the percentage of liars in the system is not specified in their model and it is given as an input. However, the mechanism is interesting since it actually defines how many agents are needed in order to have a certain degree of confidence. The system exhibits mechanisms to control the plausibility of its conclusion.

The use of likelihood theorem is a key issue also in the well-known work by Despotovic [Des04]. In his work, Despotovic describes a trust computation in a peer-to-peer environment

based solely on the information around the source entity (the trustier) and around the target (the trustee). The information around the trustier is used to estimate the trustworthiness of the network (the probability of a peer to lie when asked to report about the trustee), while the information around the trustee (i.e. agents that directly knows the trustee) is used to estimate the probability that the trustee will act in the expected way. The two estimated values are used to compute the expected trust value of the trustee applying the maximum likelihood theorem, without using any transitive path. The evaluation shows how the accuracy of this computation, based on a limited amount of information, is comparable with a computation based on all the relevant information.

Finally, in the trust model of Esfayandri [Esf01] indirect experience is based on two query protocols between agents. There are two main protocols of interaction, the exploratory protocol where the agent asks the others about known things to evaluate their degree of trust and the query protocol where the agent asks for advice from trusted agents A simple way to calculate the interaction-based trust during the exploratory stage is to compute the following formula: *number of correct replies / total number of replies.* This value is used to estimate the trustworthiness of the environment and weight any values received as recommendation.

Update recommender's trustworthiness

Few models analyzed provide a mechanism to update the trustworthiness of another agent as a recommender, i.e. the *referral* trust. The idea is to use a past experience-based approach to adjust recommended values.

An example of this mechanism is in the model by Abdul-Rahman and Hailes [Abd00] and in a similar way by Afras [Carb02]. Information coming from a recommending agent is adjusted by looking at the past recommendations from that agent and the actual outcomes that validated the interactions. For example, if A gives a recommendation to B regarding C and the transmitted value is trustworthy and B, after an interaction, judges C very trustworthy, every time B will receive a recommendation from A, it will adjust the value before using it. Note how the value is adjusted rather than discarded. It is assumed that there are differences in evaluation strategies (but the value has still some meaning) rather than maliciousness.

Time-bounded / Count-Bounded

Many reputation systems consider feedback's validity dependent on time, by using memory constraints. Two memory constraints are possible: age-bounded or count-bounded.

Early EBay implementation provided a *time-bounded* sum, so that only the ratings collected in the last 6 months are considered. In the reputation model of Sporas and Histos [Zac99] only the most recent rating between tow users is considered.

In a *count-bounded* computation an upper limit is fixed for the number of recommendation used to form a trust value. This can be due to computational constraints or efficiency, but the problem is how many feedbacks are required to keep a reliable opinion of the trustee agent.

Time Sequence

The time distribution of the ratings appears important, even if no reputation systems reviewed by the author of this work explicitly insert this parameter in the computation. The idea is simple. We should look also at the time-trend of the feedback. If two eBay sellers have the same aggregated value of +3, but obtained with these two sequences:

the second should plausibly be considered more trustworthy. The fact that the second did well the last 3 times while the first has more chance variation in its performance is an information that shouldn't be hidden.

Of course the right interpretation cannot be trivial, but it is reasonable to think that the second seller had "fixed" its problem or it has in some way evolved.

Global/Local aggregation

Another issue with ratings/feedbacks systems is whether or not feedback should be aggregated locally or globally. EBay-like systems have a global aggregation strategy, such as Epinoins.com, where users share ratings about products on the market. Massa [Mas05] pointed the problem of how global aggregation actually merges contradictions, while, as the author writes, controversial users demand local aggregation. He performed an experiment on the Epinions.com community, producing clear evidence on the loss of accuracy derived from a merely global aggregation.

Transitivity and the discounting operator

Information used in a trust computation may come from a source far from the trustee agent, propagated transitively on a trust network. Here we do not discuss if trust is transitive, but we focus on transitivity as a computational mechanism. Even form this point of view; the transmitted value should not be used straightforward.

As an example of algorithm used for propagating trust value using transitivity, we propose a simple recursive algorithm used by Golbeck [Gol02] in the context of Social Network.

If the source (trustier) has not directly rated the sink (trustee), the source queries each of its neighbors for their rating of the sink. Each neighbors' value is weighted by the trust rating the source has given to that neighbor, and the weighted average is calculated. This is shown in the following formula, where t_{ij} represents the trust from node i to node j.

$$t_{is} = \frac{\sum_{j \in adj(i)} t_{ij} t_{js}}{\sum_{j \in adj(i)} t_{ij}}$$
(A.9)

If the neighbor, node j, has directly rated the sink, it returns that rating. Otherwise, it repeats the process of querying neighbors and returns its own weighted average. This algorithm is similar to a breadth-first-search, and runs in polynomial time and it is a classical example – common to many trust solutions - of how trust can be propagated by transitivity.

The following is a list of trust models that stresses various problem related to transitivity.

In the trust model proposed by Esfayandri et al. [Esf01], authors describe how the presence of cyclic path can affect and artificially modify (for bad or for good) a trust value computed by transitivity. They consider this problem equal to a routing problem in a communication network, in order to use the same successful algorithms. In their computation they consider only paths without cycles and, to cope with the problem of contradictory values from different paths, they represent a trust value as an interval [minimum trust value, maximum trust value]. The dimension of the interval gives the uncertainty of the trust value in a similar way as [Adb00]. The authors support also contextual trust propagation, since each trust link between two nodes of the network has a "colour" representing the context, similar to the trust purpose described by Joasang.

In the model by Yu and Sing [Yus01], each agent on the transitivity path is made explicit

to the trustier. Every time an agent is asked to provide a recommendation about a third agent, it can answer with a trust value if he knows the trustee, or indicating another referral that can do the same: give a trust value or another referral. Eventually when an agent answers with a trust value and if the information is not far away to a depth limit in the chain, it can be used.

Note how the mechanism adds more information for the trustier, that can knows who the referrals are and how depth the retrieving process is taking.

Josang define a discount operation used to transitively compose beliefs. We recall how in Josang' subjective logic an opinion is a triple (beliefs, disbeliefs, uncertainty). Given a chain of two opinions (b_1,d_1,u_1) , (b_2,d_2,u_2) regarding the same issue, the discounting operator defines the resulting opinion (b,d,u) as follows:

$$b = b_1 b_2$$
 (A.10a)
 $d = b_1 d_2$ (A.10b)

$$u = d_1 + u_1 + b_1 u_2 \tag{A.10c}$$

The discounting operator has an intuitively explanation. It produces an opinion where the value of belief is the probability of both beliefs (supposed independent) to be true (20a), the value of disbelief is the value of disbelief of the second opinion under the condition that the first opinion is true. Note how the value of disbelief of the first opinion increases the final uncertainty value u as well.

Josang in [Jos05] defines also conditions for transitivity to be considered semantically meaningful. He divides trust in referral and functional trust – as already explained – and writes the following condition for a transitive trust path to be correct. A valid transitive trust path requires that the *last edge in the path represents functional trust and that all other edges in the path represent referral trust, where the functional and the referral trust edges all have the same trust purpose.*

Transitivity is therefore a powerful but high-criticized mechanism. Trust models have to consider the proximity of the information, the context, the nature of trust transferred, the propagation of uncertainty and the presence of cyclic path. Respecting these conditions increases the plausibility of the mechanism.

Social Network

Social network wants to exploit information about nodes' social relationships in order to propagate trust value. The novelty of the approach, as pointed out by Sierra in [Ope03], is in the use of sociological information as an evidence for trust. The social relationship established among agents in a multi-agent system are a simplified reflection of the more complex relationships established between human counterparts. Computational model based on social networks are largely dependent on transitivity.

Social network analysis relies on the *Small World* property that many network, notably the WWW, exhibits. Small world networks are characterized by two main properties. First, there is a high degree of *connectance* compared with random graphs. *Connectance* is the property of clustering in neighborhoods. Given a node *n*, connectance is the fraction of edges between neighbors of *n* that actually exist compared to the total number of possible edges. Small world graphs have strong connectance; neighborhoods are usually very connected. The second property, which is computationally significant, states that the average shortest path length between two nodes grows logarithmically with the size of the graph. This means that many computations can be expected to complete efficiently.

The semantic web FOAF project [FOA07] represents one of the first examples of social

network. An agent A declares a list of friends and, by transitivity, all agents that are friend of one of A's friend joins the list of A's friend as well.

Golbeck [Gol02] performed the most extensive study on Social Network, evaluated with thousands of users of the online web community *FilmTrust*. In her model trust is expressed on a discrete 10-level scale using an extension of the FOAF vocabulary to include contextual information. Trust is propagated by transitivity using a simple recursive algorithm.

Finally, in the REGRET model [Sab01], trust is based on three dimensions: the outcomes data base (ODB) based on the *direct experience* mechanism, the information data base (IDB), containing *indirect experience* information and the *sociograms* data base (SDB) that defines the agent social view of the world. Therefore sociological information about agents' relationship is made explicit in the model and represents a source of trust used along with traditional methods.

Game Theory: the utility game.

In the Game Theoretical approach, as described by Sierra and Sabater in [Sab05], trust and reputation is the result of a pragmatic game with utility functions. This approach starts from the hypothesis that agents are rational entities that chose according to the utility attached to each actions considering others' possible moves. Action could be predicted by recognizing an equilibrium to which all the agents are supposed to tend in order to maximize their collective utility.

The situation itself is another interesting parameter: its importance clearly plays a role in the balance of a trust decision. In his computational model[Mar94], Marsh considers formulas for Utility and Importance of a situation that affects the computation of trust.

Utility and Importance are values in [-1,+1] and between [0,1] respectively, subjective to each agent and dependent on the situation. Trust is computed with the formula already described:

$$T_x(y,\alpha) = U_x(\alpha) \times I_x(\alpha) \times \widehat{T_x(y)}$$
 (A.11)

Trust is therefore directly proportional to the utility and importance of the situation. Importance I is also related to the trust threshold by the following formula:

$$Threshold_x(\alpha) = I_x(\alpha) \times \frac{Perceived Risk_x(\alpha)}{Perceived Competence_x(\alpha)}$$
 (A.12)

that links the value of importance to the value of the cooperation threshold. When Importance tends to zero, according to formula 6, the level of trust tends to zero, behavior that appears anomalous. However, the cooperation threshold in the formula 21 tends also to zero, so that the agent will always trust in a situation of no importance.

As Marsh noticed, the dependencies between importance and trust is not fixed: according to Marsh's formula 20, the more the situation is important, the higher is the level of trust is needed to start a cooperation; since the action is critical only very trustworthy agents can accomplish it. However, another view is possible: the more a situation is important, a less level of trust is needed. The interpretation is the following: when an action is very important, and there are constraints such as temporal ones, the trustier accepts to cooperate even with less trustworthy agents, since it needs the situation to be done.

In this case a temporal constraint is the decisive factor that drives the trustier's decision. We note how during a trust reasoning, this evidence (the temporal constraint) makes plausible to consider the importance of the situation a factor that decreases the level of trust required. There is

therefore the notion that a plausibility test should be taken in order to understand the meaning of a generic trust mechanism, represented here by the relation between situation's importance and trust.

The Game Theoretical approach in trust can also be encoded in the design of the application. In this case, the application is designed so that trust is encoded in the equilibrium of the repeated game the agents are playing. Thus, for rational players trustworthy behavior is enforced.

[Kre82] presents the proper game-theoretic framework for analyzing reputations repeated games with incomplete information, while [Fud89] offers certain characterizations of the equilibria payoffs in the presence of reputation effects. An underlying assumption of this work is that a central trusted authority does the feedback aggregation.

An interesting example of design-based trust is described in the reputation system of Schillo et al. [Sch00]. Reputation is designed as a game played by agents in the system: the trustier, the trustee, recommending agents and observer agents that can be witnesses of the going on interaction. The environment is designed so that it is not worth it for witnesses to give false information. A witness will not say that a target agent has played dishonest in game *x* if this was not the case because the inquirer could have observed the same game and, therefore, notice that the witness is lying.

Using Risk to compute Trust

Risk is a factor that cannot be neglected in trust computations. The secure trust engine considers risk analysis an essential part of the final trust-based decision. A high-level view of Secure is depicted in figure A.4. A trust-based decision is taken when a trustee entity puts a request. The decision making process is based on a trust value computation and a risk analysis, based on the evidences stored in the evidence manager, accumulated in all the previous interactions or received by other peers.

www.zeallsoft.com

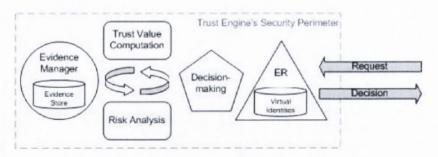


Figure A.4 High Level Architecture of the SECURE trust engine

Secure's risk module can dynamically evaluate the risk involved in the interaction and its decision-making policy is to choose (or suggest to the user) the action that would maintain the appropriate cost/benefit [Cah06].

Marsh gives emphasis to the role of risk in his trust model. As we have already seen in formula A.12, the cooperation threshold, that defines the minimum amount of trust needed to start an interaction, is directly dependent on the perceived risk.

In his work, a clear difference is made between the three concepts of *risk*, *ignorance* and *uncertainty*. In situation of *risk*, it is still possible to estimate the likelihood of events, in

uncertain situations these probabilities are not known, while due to *ignorance* we could not known which probabilities should be computed. When an agent does not know anything about another, it is in a situation of ignorance. The agent will use an initial trust value according to past experiences similar to that situation or following its policy for new entities.

In the case where risk can be computed, any risk assessment can be used – for example, the one proposed by Secure based on a classical benefit/loss analysis – and, once a usable metric have been defined, values can be incorporated in a trust-based decision.

Analysis how risk is computed is outside the scope of this thesis. We definitely consider risk an argument in a trust-based decision, even if we keep it separated and we treat it as a datum. In line with many authors, we consider risk a more defensive concept that suits mainly security-based scenarios, focused on assessing the likelihood of events and their attached benefits/loss, while trust is a more cooperation-oriented concept.

If trustee A and B have a comparable risk, a trustier agent will select the one he trust more and vice versa. Decisions can be made only on one of the two factors, but relying on both make the process more plausible and complete.

In our model, risk affects a trust-based opinion rather than be part of the actual trust-based computation. Risk therefore can revert a trust-decision, but it does not affect the trustworthiness of agents. It is an exogenous factor that logically is separated by trust, but essential in a decision. This decision is in line with our intention of being focused around the peculiar nature of trust.

Cognitive models, the Falcone-Castelfranchi model

The trust model proposed by Castelfranchi and Falcone [Cas05] is a clear example of a cognitive trust model. The basis of their model is the strong relation between trust and delegation. They claim that 'trust is the mental background of delegation' [Cas05]. In other words, the decision that takes an agent x to delegate a task to agent y is based on a specific set of beliefs and goals and this mental state is what we call trust. Therefore, 'only an agent with goals and beliefs can trust'.

To build a mental state of trust, the basic beliefs that an agent needs are:

- Competence belief: the agent should believe that y can actually do the task.
- Dependence belief: the agent believes that y is necessary to perform the task or that it is better to rely on y to do it.
- Disposition belief: not only is necessary that y could do the task, but that it will actually
 do the task. In case of an intentional agent, the disposition belief must be articulated in
 and supported by two more beliefs:
 - Willingness belief: the agent believes that y has decided and intends to do α (where α is the action that allows the goal g.
 - Persistence belief: the agent believes that y is stable in its intentions of doing α .

Supported and implied by the previous beliefs, another belief arises:

- Fulfillment belief: if the agent "trust in y for g", the agent decides:
 - (i) not renouncing to goal g,
 - (ii) not personally bringing it about,
 - (iii) not searching for alternatives to y, and
 - (iv) to pursue g through v.

To summarize, trust is a set of mental attitudes characterizing the delegating agent's mind (x) which prefers another agent (y) doing the action. Y is a cognitive agent, so x believes that y intends to do the action and y will persist in this.

The authors propose a partial computational approach to their model in [Cas00], where they implement a hierarchical fuzzy network to propagate trust values. They present an interesting methodology to compute trust. They identified four macro-areas: Similarity & Categorization, Reasoning, Direct Experience and Indirect Experience.

For each area pieces of evidence are collected, and the resulting trust value is computed by taking into consideration: the strength/reliability of the source used to collect the evidence, the content of the belief collected, who/what the source is, how this source evaluates the belief, how the trustier evaluates this source. As the authors wrote: the interesting thing [...] is that starting from finding the sources of trust we are obliged to consider the trustworthiness of these sources. [Cas03]

Interesting for our purpose is therefore the notion of computing trust values according to the plausibility of the evidence source. The work is highly pertinent since the notion of plausibility of sources is central in evaluating trust. Moreover, the rich nature of trust presented is in accordance with our vision, and the trust evidence suggested by the four macro-areas inform the definition of our defeasible trust scheme.

As many authors pointed out, Castelfranchi's cognitive model does not have an efficient and complete computational version. The example that authors give in [Cas00] is limited and the trust values for each macro-area is given as an input, and the computation performed in the paper is reduced to a strategy to combine the different source of information using a fuzzy-network approach.

Trust-Based Heuristics: two examples

Trust can be the result of a computation over elements of the application under analysis by mean of some kind of heuristics or intuitions. We specifically focus on computations based on domain elements, since any kind of other evidence, from human feedback to digital certificates, can potentially be manipulated by a heuristic.

We consider two meaningful example of domain elements-based heuristic: the *PageRank* [Pag99] metric and one of its versions in the context of the *Wikipedia Project*.

Many authors, notably Massa [Mas03] identified in Google *PageRank* all the elements of a trust metric. In this specific case, the application is the whole Web, seen as an interconnection of mutually linked web sites, where *PageRank* selects as evidences the outgoing and ingoing links of a web page. This is an element of the application part of its core functionalities, not implemented for explicitly supporting trust. The computation described in the *PageRank* formula is performed over these evidences to derive a value for a specific web page or related web sites.

The second example, closely related, is represented by the work that McGuiness [McG06] did on the Wikipedia project. In order to assess the trustworthiness of a Wikipedia article, the author applied heuristics based on a version of the PageRank algorithm. They considered the relative number of times an article's name appears as a link or as plain text in the application. The relative amount of time an article was a link was used as a trust metric. The metric proposed was the following

$$T_{doc} = \frac{occ([[d]])}{occ([[d]]) + occ(d)}$$
 (A.13)

The meaning of the factors is the following:

- T_{doc} is the trust value associated with a document
- d is the title of the articles, such as "Beer" or "Theory of Relativity"
- Occ(d) is the number of occurrences of the article title d in the whole Wikipedia

• Occ([d]) is the occurrence of the article's title as a link to the article d in the whole Wikipedia

Discussion

As discussed in chapter 2, the weak point of a trust-based heuristics is its justifications and lack of systematicity.

The justification for selecting elements as trust evidences in *PageRank* could be sustained noting that the act of linking a page is like an implicit act of recommending that page or recognizing its reputation [Pag99]. In other words, the justification is that a generic mechanism of trust (the *indirect experience*) has an implicit correspondence in the dynamic of the application. Note how the justification/interpretation is a presumption, not a valid statement. In the context of the WWW the hypothesis appears more justified than in Wikipedia. While *PageRank* has been an effective way to rank pages, its applicability to *Wikipedia* can be severally argued: an expert Wikipedia's user may argue that in Wikipedia there are automatic procedures that link articles, or that an author may link articles independently by the content of the linked article (for example for the sake of completeness). These examples, as many others, shows how the selected heuristics it cannot be applied straightforward without a critical analysis of its applicability and its relevance to the context.

Monitoring the system: the model by Carter

As a good example of monitoring approaches, we analyse the trust model proposed by Carter. The main idea behind the reputation model presented by Carter et al. [Car02] is that 'the reputation of an agent is based on the degree of fulfillment of roles ascribed to it by the society'. If the society judges that agents have met their roles, they are rewarded with a positive reputation; otherwise they are punished with a negative reputation.

'Each society has its own set of roles. As such, the reputation ascribed as a result of these roles only makes sense in the context of that particular society'. According to Carter, it is impossible to universalize the calculation of reputation.

The author formalizes the set of roles within an information-sharing society and proposes methods to calculate the degree of satisfaction of each roles. An information-sharing society is a society of agents that attempt to exchange relevant information with each other in the hope of satisfying a user's request. He identifies five roles:

- Social information provider: 'Users of the society should regularly contribute new knowledge about their friends to the society'. This role exemplifies the degree of connectivity of an agent with its community. The degree to which the social information provider role is satisfied by a given user is calculated as the summation of all its recommendations, mapped in the interval [0,1].
- Interactivity role: 'Users are expected to regularly use the system'. Without this participation
 the system becomes useless. The degree of satisfaction for this role is calculated as the
 number of user operations during a certain period of time divided by the total number of
 operations performed by all the users in the system during the same period.
- Content provider: 'Users should provide the society with knowledge objects that reflect their
 own areas of expertise'. The idea is that users that create information related to their areas of
 expertise will produce higher quality content related to their interest than those who do not.
- Administrative feedback role: 'Users are expected to provide feedback information on the quality of the system. These qualities include easy-of-use, speed, stability, and quality of information'. Users are said to satisfy this role by providing such information.

Longevity role: 'Users should be encouraged to maintain a high reputation to promote the longevity of the system'. The degree of satisfaction of this role is measured taking into account the average reputation of the user in the community.

Given that, the user's overall reputation is calculated as a weighted aggregation of the degree of fulfillment of each role. The weights are entirely dependent on the specific society. The reputation value for each agent is calculated by a centralized mechanism that monitors the system

In Carter's method trust is deduced directly by monitoring the environment and identifying elements that could be useful in the computation. Elements are selected by mean of an explicit notion of trust, represented in Carter by the five expected roles that an agent should fulfill. The five roles suggest which elements of the domain should be used: any element revealing that the agent is fulfilling or not these roles is trust evidence. The model is therefore justified process.

Miscellanea

The problem of the starting value: disposition, attitude, constraints.

In any computational trust mechanism, the problem of the initial trust value is of great importance. Even solutions that propagate trust values have strategies to define initial values.

A common point is that an entity, in absence of previous interactions or evidence regarding a new trustee, uses a *dispositional trust value*, a general disposition of the agent in its environment. Marsh called it *basic trust*, that depends on all the information an agent holds about the environment, encompassing past interactions coupled with similarity among situations and agents, or it might depend simply on an *a priori* disposition.

Agents can have an *optimistic* or *pessimistic disposition* about others, and different threshold value for starting an interaction. In Marsh optimistic agents always select the maximum trust value from the range of experiences they had and vice versa.

In the *CertainTrust* model, *pessimistic* agents assign a trust value of 0 to unknown agent, meaning that trust can be computed only after having collected some evidence. This reflects an attitude "I believe entities are untrustworthy, unless I know the opposite with high certainty". *Moderate* agents assign a starting value of 0.5 to new agent (on a scale [0..1]) while optimistic agents accept any interactions with a new agent, reflecting a position where entities are considered trustworthy unless opposite evidence is discovered.

Trust models that manage uncertainty and represent trust value as a probabilistic function implement generally the following solution. When nothing is known about another agent, trust value is a uniform distribution: no value is more likely than another. For instance, in the STRUDEL model by Quercia [Que06], trust is represented by a probability distribution over a discrete n-level set. As for bootstrapping, when the peer p_x meets p_y for the first time, it has no information about p_y ; its beliefs are thus uniformly distributed with a value 1/n.

Finally, an initial trust value could be defined using constraints. A new peer, or new recommender, is accepted by the trustier only if it can satisfy specific constraints. On the contrary, when values are accepted without constraints we have a *blind trust formation*. A constraint can be described as a policy. For example, NeuroGrid [Jose02] trust model allows to define constraints on the length of a transitive chain (a concept similar to a Time-to-live for network packet), where new peers are not trusted if they are at a further distance. The model by Sierra [Ope03] defines an initial trust value dependant on the distance of the peer, decreasing exponentially as shown in table A.1

Table A.1 Initial Trust value assignment in Sierra.

| Distance (hop) | 1 | 2 | 3 | 4 | >5 |
|---------------------|-----|----|----|----|----|
| Initial Trust Value | 100 | 50 | 30 | 20 | 0 |

We note how these constraints are actually tests over the validity of a mechanism – in this case transitivity – that give clues about the defeasibility of the mechanism.

Policymaker [Bla96] allows defining condition to accept new trustee in the context of PKXI authentication such as a valid CA within a certain name space, an email address belonging to certain domains etc...

BBK [Bet94] allows defining a flexible set of predicates that a new entity must satisfy in order to start interactions, based on a subset of properties that an entity can show in a specific environment.

Constraints are actually rules not inherent to the computational trust mechanism used but more oriented to add an extra-security layer.

Finally, in the model by Garg [Gar06], a new entity is trusted only if an old member of the community guarantees for it. The guarantor risks some of his reputation if the new entity acts badly. This idea hides actually a generic pattern of reasoning: entity X can trust Y since Z, the guarantor, has something to lose. We could classify this reasoning in the game-theoretical approach and it represents a generic trust pattern we will use in the definition of our model.

Once the starting value is defined, the value evolves using one of the computational mechanisms described earlier.

System Trust

A large majority of social scientists identify three type of trust: dispositional, situational, and system trust. Dispositional and situational trust has been already described in this section. System Trust is a property of the system, rather than an entity personal attitude (dispositional trust) or a situation-based decision (situational trust). System trust is defined by McKnight [McK00] as follows: "the extent to which an entity believes that proper system structures are in place to enable it to anticipate a successful future endeavor". System trust is therefore the confidence that safeguards, regulations, laws, guarantee and security mechanisms embedded in the system reduce the potential negative consequences of trusting behavior. Agents are trusted because of the systems they are interacting in independently from their characteristics. Usually system trust can be used in situation where little of nothing is known about an agent but we trust the system as a guarantor.

System trust can be encoded in the design of the system making it partially overlapping Game Theoretical approaches.

We note how system trust is a real type of trust. Agents actually trust that the rules encoded in the system will be enforced, but no total guarantee is given about this. Castelfranchi and Falcone explained in [Cas00] how the presence of a contract between agents does not exclude the presence of trust and risk..

Appendix B Statistical Tools

In this section we describe basic computational tools that we will use in the rest of the thesis and in the definition of the trust schemes. The computations performed are mainly descriptive statistical methods for ranking and analyse data distributions.

Basic Statistical Operators

Probability Distribution Function (pdf)

A probability (*density*) distribution is any function f(x) with the following property

$$\forall x \in [-\infty, \infty], f(x) \ge 0$$
 (B.1)

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$
 (B.2)

X is called a continuous random variable. The typical *pdf* we analyse represents the distribution of entities according to the value of some specific indicators linked to a trust scheme.

Cumulative distribution function (pdf)

Given a pdf f(x), its cumulative distribution function is:

$$F(x) = \int_{-\infty}^{x} f(x) dx$$
 (B.3)

The two functions have the following interpretation. The value of F in a given point a is the probability that the random variable x is lower then a.

Graphically, this value is the area below the pdf f(x) in $[-\infty, \mathbf{a}]$. The probability distribution of a random variable is often characterised by a small number of parameters, which also have a practical interpretation. These synthetic indicators are essential for the computation of many trust schemes.

The moments of a random variable are the essential indicators. We define the n^{th} momentum M about c of the random variable x the quantity:

$$M_n(x) = \int_{-\infty}^{x} (x - c)^n f(x) dx \tag{B.4}$$

Momentums represent a set of values that can entirely describe the characteristics of a pdf. The I^{st} momentum about zero is called the expected value of the distribution that is therefore:

$$M_0(x) = E(x) = \mu = \int_{-\infty}^{x} x f(x) dx$$
 (B.5)

And it represents, for a continuous variable, the average value. The moments about 0 is called the centred momentum and describe the distribution independently from its average value. The 2^{nd} momentum about μ is called standard deviation σ :

$$\sigma = \int_{-\infty}^{x} (x - \mu)^2 f(x) dx \tag{B.6}$$

The standard deviation is an indicator of the compactness of the distribution. A large standard deviation indicates that the data points are far from the mean and a small standard deviation indicates that they are clustered closely around the mean. Outliers' values, largely far from the mean, are therefore more serious when σ is small, while there are more plausible when σ is big.

Mode

The mode of a pdf f(x) is the most frequently occurred value, therefore:

$$d: \forall x, f(d) > f(x)$$
 (B.7)

Note that the mode d is in general not univocal.

Median

The median of a random variable x with pdf f(x) is a point m so that:

$$F(m) = \int_{-\infty}^{m} f(x)dx = F(m) = \frac{1}{2}$$
 (B.8)

The interpretation of the median value is the following: exactly half of the population has a value lower than m and half a higher value. Note that m and μ are the same when f(x) is symmetric, but they are usually different. They are related by the following relation:

$$|\mu - m| < \sigma \tag{B.9}$$

Therefore the two value can differs at least from a standard deviation. An interesting property of the median is that it is totally insensitive to "outliers" (such as occasional, rare, false experimental readings). The mode is very robust in the presence of outliers, while the median is rather sensitive. This because mode and median do not take in consideration where data are located but rather their value.

Percentile

A percentile is the value of a random variable below which a certain percent of observations fall. For example, a 20-percentile is the value a so that F(a) = 0.2. The 50-percentile is the median.

In a distribution of a population according to some indicators, percentile has a useful interpretation. If we are interested in that value that only 10% of the individuals has larger, this information is the 90-Percentile.

Statistical Tests and Concepts

Statistical significance tests

A statistical significance tests is used to understand if there is a statistically significant difference between two distributions of data, or these differences can be explained as effect of chance variation. A confidence level is set to decide the tolerable amount of uncertainty of the test. The hypothesis to test is called the *null hypothesis*, and involves usually the comparison of two average values coming from the two set of data. Our methods will often imply the comparison of two set of results coming from different computations performed over the same population of individuals, typically the output of our method and the output of an another independent trust metrics used as comparison.

Therefore, we use significance test to understand if the two trust metrics can be considered identical or there is a significant difference. The statistical test will be a paired t-test, used to test the two different set of a data where it is possible to identify a bi-univocal relation among the individuals of the two groups, in our case the entities composing the population. Where this relation cannot be defined – such as when the population is composed by anonymous entity – an unpaired t-test can be used. Since in a paired t-test there is less uncertainty – since individuals can be matched – given two set of results, a paired t-test requires less difference among the data than an unpaired one.

Since our computation is based on rankings, statistical tests will be performed with the more specific Spearman's rank correlation test (ρ) , that is a value in [-1,1] that measures the degree of similarity among two rankings. The formula is the following:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$
 (B.10)

As correlation between two variables, a value of 0 means no similarity between the two rankings, 1 maximum similarity and -1 means that the two rankings are opposite.

Sampling

Sampling consists in analysing a set of individuals belonging to a population on order to yield some knowledge of the entire population. Since the number of individuals analysed is smaller than the entire population, the main advantage of sampling is the time required for the analysis. In some case, a comprehensive observation of all the individuals – due to the high number or the impossibility of accessing all of them - is the only possible techniques.

Of course, conclusions based on sampling are affected by uncertainty, since they are not derived from the entire population. This uncertainty is a function of n, the size of the sampling test: lower value of n leads to high uncertainty and vice versa.

Sampling is described by the following tow main results. Given a pdf f(x), with μ and σ , if we collect a sample of size n, μ_n and σ_n are generally different form μ and σ . If we collect different sampling sets all of size n, and we build the distribution of μ_n , we obtain a Gaussian distribution with expected value equals to μ and standard deviation equals to $\frac{\sigma}{\sqrt{n}}$.

Given a certain level of uncertainty U (expressed as a percentage) and an error size E, the problem is therefore to choose the correct size n so that the sampled value differs from the actual less than E with a level of uncertainty U. The formula used to determine the sample size n is equal to:

$$n = \left(\frac{z_{crit}\sigma}{E}\right)^2 \tag{B.11}$$

Where:

 σ = standard deviation of the entire population

E = error' size we are keen to tolerate. The real μ has to be in the confidence interval $[\mu_n$ -E, μ_n +E]

 Z_{crit} = the critical value corresponding to the certainty level U fixed. If U is equal to 0.95, Z_{crit} is equal to 1.98 and μ will be 95% of the time in the interval [μ_n -E, μ_n +E]

When we have to estimate proportion of a population – for example, the percentage of population above a certain value – the following formula can be used:

$$n = p(1-p)^{\frac{\pi_{crit}}{E}}$$
 (B.12)

Where p is the percentage of the sampled population.

Correlation/Relation among variable

The correlation coefficient is defined as follows:

$$C_{xy} = \frac{E((X - \mu_X)(Y - \mu_y)}{\sigma_X \sigma_y}$$
 (B.13)

 C_{xy} indicates the strength and direction of a linear relationship between two random variables. In general statistical usage, *correlation* or co-relation refers to the departure of two variables from independence. Correlation is useful to estimate if two random variables may have a dependency or a similar behaviour.

Order and Ranking Method

Order and Ranking statistics is a branch of non-parametric statistics, that concerns with the analysis of data using *distribution free* methods as they do not rely on assumptions that the data are drawn from a given probability distribution. The set of population we are investigating in our trust computations is clearly distribution-free, since we cannot assume any a priori probability distribution able to fit population's trust values. Therefore Order and Ranking statistics offer adequate computational methods to analyse our set of values.

In order statistics, great importance has special values in the set of data like maximum, minimum, median, percentile. Ranking statistics analyses and make interpretation of set of data by ranking individuals according to a criteria for which a total order exists – i.e. every element is greater or less than any other one.

The ranking method is the base of our computational models. In its simplest purely aggregated form, each trust scheme generates a ranking among entities, and the final trust value is a ranking obtained by summing all the different rankings. Value's strength is also quantified by comparing them with population mean, or using percentile.

Types of ranking. Different rankings are defined on the basis of how treat equal measurements:

- 1. Standard competition ranking, such as 1224
- 2. Modified competition ranking, such as 1334
- 3. Dense ranking, such as 1223
- 4. Ordinal ranking, such as 1234, where equal values are randomly or arbitrarily ordered.
- 5. Fractional ranking, such as 1,2.5,2.5, 4

We use the *fractional ranking*, due to its properties that guarantee more fairness among the individuals, especially when different rankings have to be sum to form a global one. The *dense ranking* introduces a bias among different rankings, since the ranking does not reflect the actual position of the individual in the ranking that depends on how dense the ranking is. The *ordinal ranking* introduces an arbitrary factor, while the second one is biased for entities with not univocal values.

After having ranked a set of data, we study the distribution obtained by considering the fundamental measurements defined earlier.

Percentile ranking is useful to understand the position of an individual among the population. For example, if an individual show a value of a, and a is the 95-percentile value p_{95} , that means that only 5 per cent of the population has a value higher than it. This information cannot be deducted by comparing the value a with μ or σ , since they can only quantify how far is the value from the average, but not the distribution of the individuals.

We introduce another useful indicator on a ranking distribution that considers the impact of a particular group of individual on the global set of value. Given a set I of individuals i, and an associated normalized value v(i), we define R(n): $n \rightarrow v(n)$ the (descending) ranked distribution of the values v(i), that for each position n of the rank returns the value v(n) of the individual at position. Note that this distribution shows the values v(i) as a function of the position n in the ranking. An example is depicted in picture 4.2. The nth-contribution is defined as follows:

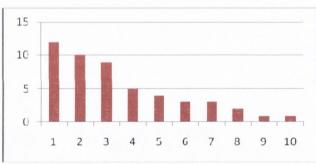


Figure B.1 The function R(n)

We define the n^{th} -contribution as follows

$$C(n) = \int_1^n R(n)$$
 (B.14)

The n^{th} -contribution is therefore a measure of how much the first n individuals contributed to the ranking. For example, if C(5) is equal to 0.5, this means that half of the contributions were done by the first 5 individuals. The n^{th} contribution can be computed also over percentage of rank, for example C(5%) is the contribution of the top 5% ranked individual. Note the difference with the n-percentile rank, which is computed over the frequency distribution of values v(i), and returns a value v(a) below which n% of the population values fall. Percentile gives information about how values are distributed, but it does not give any information about the impact of an interval of value.

Confounding Variables

A confounding variable (also confounding factor, lurking variable, a confound, or confounder) is an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable.

We need to control for these factors to avoid what is known as a type 1 error: A 'false positive' conclusion that the dependent variables are in a causal relationship with the independent variable. It is of extreme importance to test each evidence or trust scheme avoiding that results are not plausible for the effect of an external variable that represent the dominant factors, invalidating all the computation and leading to invalid results.

As we see in the next chapter, many critical tests on trust schemes investigate the potential effect of an external variable - maybe part of another trust scheme – that could determine the values of the first trust scheme. We have to avoid situations in which the several evidence is used multiple time when in reality they all refers to a same phenomena.

In order to remove the effect of a confounding variable, several techniques are possible.

Case-control studies assign confounders to both groups, cases and controls, equally. For example if somebody wanted to study the cause of myocardial infarct and thinks that the age is a probable confounding variable, each 67 years old infarct patient will be matched with a healthy 67 year old "control" person. In case-control studies, matched variables most often are the age and sex.

Cohort studies: A degree of matching is also possible and it is often done by only admitting certain age groups or a certain sex into the study population, and thus all cohorts are comparable in regard to the possible confounding variable. For example, if age and sex are thought to be confounders, only 40 to 50 years old males would be involved in a cohort study that would assess the myocardial infarct risk in cohorts that either are physically active or inactive.

Stratification: As in the example above, physical activity is thought to be a behaviour that protects from myocardial infarct; and age is assumed to be a possible confounder. The data sampled is then stratified by age group – this means, the association between activity and infarct would be analyzed per each age group. If the different age groups (or age strata) yield much different risk ratios, age must be viewed as a confounding variable.

One major problem is that confounding variables are not always known or measurable. This leads to 'residual confounding'. Hence, randomization is often the best solution as, if performed successfully on sufficiently large numbers, all confounding variables (known and unknown) will be equally distributed across all study groups.

Appendix C Trust Schemes and Critical Questions: a detailed discussion

Introduction

In this appendix we provide a more detailed description of our set of defeasible trust schemes forming the content of our trust-based reasoning.

As described in chapter III, a trust scheme is an *argumentation scheme* specific for trust. Therefore, a trust scheme is a generic defeasible reason to support trust or distrust. This section described the trust schema we identified, divided in 8 overlapping macro-areas:

- 8. Time-based trust scheme
- 9. Trust-Scheme based on Information Sharing
- 10. Trust scheme linked to social role
- 11. Trust scheme based on activity analysis
- 12. Trust scheme based on outcomes
- 13. Trust scheme based on statistics and grouping
- 14. Trust scheme based on Game theory and Cognitive models
- 15. Trust scheme based on factors exogenous to trust, but still related to it

The source of our trust scheme is represented by the state-of-the art of computational trust models and social science literature. Our work is to extract, from the state-of-the-art, the underlying reasons that can form a scheme. Some trust schemes haven been already proven to be effective-at least partially - in some trust computations. Some other trust schemes we identified represent a novelty: they are still grounded in social science but the do not have any no computational model or evaluation regarding their validity.

Another novelty is represented by the investigation of each trust scheme defeasible nature, essential to support our argumentation.

Since trust scheme are presumptions, they are not definitive evidence for trusting an entity when they are positive, and vice versa when they are negative. Anyway, when a set of plausible trust schemes is properly aggregated, they can build a presumption stronger enough for the trustier to take a decision, as our evaluation shows. Trust schemes consider the same aspect from different angles, and their application helps practitioners to better design an effective trust computation.

Our list does not claim to be comprehensive, but it claims to represent a set solid enough to support an useful trust computation.

How trust schemes are described

In chapter 4 we provide a formal definition of trust scheme as an inference graph with specific elements and constraints. In the rest of the discussion, for clarity's sake, we preferred to present trust schemes in an informal way, providing for some of them an inference graph as example. Nevertheless, our description of each trust scheme is still modelled after our formal definition and it always encompasses:

- 1. a description of each scheme's computation, or set of possible computations
- 2. each scheme defeasible assumptions

3. the set of crucial questions attached to each scheme, i.e. a list of potential defeaters and supporters

Time-based Trust Schemes

This class of schemes build trust arguments using only information about time, usually temporal intervals between interactions or interactions' timestamp.

They do not consider what was done during an interaction and - more important - how it has been done. The focus is about when it happened.

They represent a contribution to trust studies, since no evaluation of their effectiveness has been performed, while they are present in some high-level computational model of trust and their importance is largely acknowledged both by social studies and by common sense.

Time-based trust schemes are: longevity, persistency, regularity and stability. Stability has been placed among time-based trust schemes since it focuses on how entities' properties changes over time.

The relationship between time and trust has been always considered extremely strong. Even in common sense, we all know sentences such as "I have been doing this job since 30 years" or "Car makers since 1902". Time is "sold" as an evidence of the trustworthiness and reliability of an entity. Concepts such as experience, ability to adapt and survive, resistance and reputation are all linked to time and represent clear trust evidence. For instance, Carter includes in his model a longevity factor [Car02]. We now describe the trust schemes identified.

Longevity

The presumption behind this scheme is: I trust an entity because of its longevity in the environment. The scheme produces a trust argument for older entities, while it represents a warning for very young entity. The underlying reason is that longevity supports two ingredients of trust: entity's experience and reliability, derived from the ability to survive that an entity shows with its longevity.

The basic computation of the scheme implies simply to evaluate a time interval between the time t_0 – the birth of the entities - and t_{last} , the timestamp of the last useful moment.

Longevity is a presumption and several tests can increase or decrease its plausibility. The following are the critical question identified many of them applicable to all the time-based schemes:

Is the choice of t_0 *and* t_{last} *correct?*

How you decide that an entity is part of the environment or no longer in the environment?

Usually, t_0 is the time of birth of the entity, and t_{last} is the present time. An alternative choice for t_0 could be the time of the first interaction, and t_{last} the one of the last interaction, or the time of the "dismissing" of the entity, if it is known that the entity is no longer in the environment. The interval defined in this way could be a better estimation of the age of the entity in relation to its activity. A typical example is a member registered in an online forum: t_0 could be the time of registration, or the time of the first message sent, while t_0 could be the present time or the time of the last message.

The uncertainty of the trust scheme is linked to the uncertainty of t_0 and t_{last} , derived form the source of information used. Extra question to be considered are:

How you decide that an entity is part of the environment or no longer in the environment? Is longevity linked to a specific action or more action/context?

This means that we could decide to consider an entity active in the environment if its actions are at least of a certain type. If an entity performs a minor or negligible interaction, this could not be considered enough to consider the entity alive. We note how these critical questions imply focusing the attention on what the entity did – type of interaction – and not only on when it was done. The question actually introduces a first example of logical relationship between the two macro-areas activity and time.

As an example, an author of Wikipedia could edit an article by adding a paragraph, uploading a picture or only by correcting a minor grammatical error.

Is the environment competitive and selective?

Longevity is a much stronger argument if the environment entities are interacting is competitive. If this hypothesis is true, a high longevity is presumably a sign that the entity is able to survive and compete with the others, able to provide a service and quality of interactions that are enough for the entity to stay in the environment. This presumption can become stronger when considering the following question:

What is the cost of staying alive?

If entities can stay in the environment at no cost, longevity is a weaker argument. The question undercuts the link between longevity and reliability and ability to survive. On the contrary, if staying in the environment requires an effort for the entity, longevity becomes a stronger evidence of a deeper relation between entity and environment. As an example, if an entity is subscribed to a free service and to a payment one, if the two services do not satisfy it anymore, it is more likely that it will unsubscribe to the second one that is costing money, than the first one, that is free. Or, if the services are similar, there is no reason why the entity should not switch to the free one. Thus, longevity is a stronger argument for the first service.

Another example is virtual online entities. If the registration, as usual, is a free procedure, entities can be created at any time without limitation and a physical entity may be linked to several virtual identities as investigated by Friedman and al. [Fri99]

In this context, a virtual identity with higher longevity is still preferred to an entity with a lower longevity. Of course, the relation with other trust schemes will increase or decrease the global trustworthiness of the entity.

Is there a positive or negative relationship between time and entities abilities?

The critical question considers if an old entity has less ability to deliver a good outcome since some of its properties have a negative relationship with time.

Note how the scheme can have a different effect.

It could also be the case the trust scheme is a presumption that is completely wrong, meaning that in the context the opposite is true: entity with low longevity are better. For example, a 20-years old person is better than a 60-years old person in running the 100 mt. Again, the critical question has to detect it. In this case we have to investigate if enteritis in the environment has some abilities that decrease with time, so that it makes inapplicable the scheme.

If we want to depict the scheme on an inference graph, we could obtain a situation similar to the one showed in figure C.1, showing the generic arguments of the scheme and a potential situation of an argumentation of a trustier agent in a specific epistemological state regarding the trustworthiness of a member of an online forum.

Trust Scheme Persistency/Regularity

According to this trust schemes, entities that persistently and regularly interact in the environment should be trusted, while low persistency and regularity should be investigate since it is a potential evidence for distrusting.

The two trust schemes are closely linked, and they are actually two aspects of the same computation.

Persistency quantifies if an entity show signs of activity in the environment every fixed interval of time ΔT . We called ΔT the persistency time interval (symbol Π). An entity is regular if the time interval between two consecutive interactions is relatively constant and not subject to high variance.

Both the trust scheme are based on time interval between interactions, but while persistency is based on a time interval decided externally, regularity is based on the distribution of time intervals between actions.

This means that, if an entity interacts every 20 days, we could compute a persistency based on an interval of 1 week and find that an entity has low persistency but still good regularity. For stronger evidence, both of the information should be considered. The computation attached to this trust scheme is now described.

When an entity is persistent and regular, this means that is likely to be present in the environment when needed; it is a sign of transparency and it guarantees that the entity is well-known. Therefore, there is a direct link with the trust ingredient accountability and accessibility. Persistency and regularity are also a sign of a special and recurrent relationship between the entity and its environment, showing how the entity is familiar and bonded to the community, accomplishing its social role in respect to other community peers, to use the a concept familiar to Carter's trust model. Defeasibly, regular presence in the environment increases the perceived trustworthiness and reliability of a trustee entity.

Persistency Formula

The formal computation can be written as follows. Given:

 Π = time interval

T = activity threshold

 T_0 = starting time

 $T_{last} = ending time$

$$N_{tot} = total \ number \ of \ intervals = \frac{t_{last} - t_0}{\Pi}$$

We define a function A(t) that has value 1 if at time t an interaction above the threshold happened and 0 if not. In the simplest form A(t) checks if an interaction occurred (the threshold is set to 0). In more elaborated form, the threshold could be set to a non-null value: the type and complexity of the interaction could be considered as discussed before for longevity.

We define a function I (interval function) defined over a time interval as follows

$$I(t_1, t_2) = \begin{cases} 1 & \text{if } \int_{t1}^{t2} A(t) > 0 \\ 0 & \text{in other cases} \end{cases}$$

The persistency can be defined as follows

$$P = \frac{\sum_{n=0}^{N\text{tot}} I(nt, (n+1)t)}{N_{\text{tot}}}$$

In our definition of persistency we divide the total time in intervals of size Π . Persistence is the percentage of intervals where the entity has been enough active, i.e. above the threshold.

In this definition of persistency some elements are missing: it would be more complete to consider also the maximum amount of interval where the entity was active or not, since this information could be used for a better understanding of the plausibility. Usually, by performing a more computation varying the size of Π , this information can be appreciated.

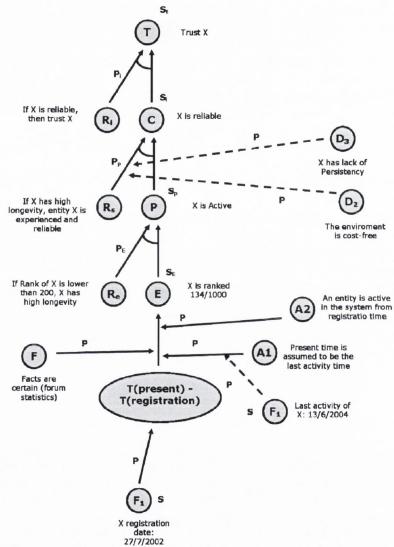


Figure C.1 Longevity Inference Graph at a certain epistemological state. (Example taken from experiment I)

Regularity Formula

The formula for computing regularity considers the time between two consecutive interactions and build the distribution D(t) of such intervals.

Regularity is high when the average of the distribution is low – meaning that the member usually has a short time before two interactions - and the standard deviation is low - meaning that the entity does not have a high variation in its time interval. A high standard deviation will denote

that the entity could have high period of non-interactions affecting the regularity. The *max* and *min* value of the function should be considered as well for complete the analysis.

As stated before, the two computations are slightly different: in the first one the length of the period of time Π is decided externally while regularity is a property inherent to the distribution.

High value for both persistency and regularity means that the entity is available in the environment constantly and with high frequency.

Regarding the critical tests to be performed and the assumptions to be considered, the following tests described for longevity are still valid:

How activity is defined? How is the threshold settled? Can you quantify it?

The question considers the threshold to consider the interaction of an entity. The threshold could be a function of the number of interactions in a time interval, and the interaction can be defined as complex as needed.

Other critical questions involve the investigation of the time interval:

Is the time interval correct?

This critical questions focus on the choice of Π . This parameter should not be taken too high, so that all the entities result persistent, or too small to have the opposite effect. Regularity could suffer the same problem, so that it became hard to make the correct interpretation of the results. Increasing the complexity of the interaction required to trigger persistence and reducing Π make conditions for persistency and regularity harder to satisfy. This can help in selecting the value of the parameters, in situation where starting values of Π did not reveal any substantial differences among entities.

When persistency and regularity are computed among a group of entities – as in our evaluations – it is still possible to make relative comparisons of persistency and regularity values independently form the choice of the time interval. Nevertheless, the correct choice of Π is essential in order not to generate true negative or false negative. The following critical tests can help:

Is the interaction supposed to be persistent and regular?

If the interaction we are monitoring is not supposed to be regular or persistent, the trust scheme should be defeated. Activity could be linked to external constraints and not been a free choice of the entity, so that assessing regularity and persistency make no sense.

The presence of cycles in the interactions should also be monitored. It could be the case that the interactions are not supposed to be persistent by its nature. As an example, the action of buying shares on the stock market is possible only 5 days a week, therefore by choosing Π equal to one day, we should consider that 2 days every 5 could not register any activity, and this is not a lack of regularity or persistency. The same could be for period of holiday or forced interruption of service.

Is the action frequent enough to collect many data?

If the type of interaction we are interested in has a very low frequency, the trust scheme may have lost a lot of meaning. Some actions could be very rare or once in a lifetime, therefore the trust scheme could not be applicable. All depends on the frequency of the interactions. An heuristic rules could be to choice Π as a function to the frequency f_a of the activity, avoiding situation in which $f_a \gg \Pi$ or $f_a \ll \Pi$. When possible, a statistical analysis of the entity communities lead to better and not easy-to-defeat estimation of f_a and therefore Π .

Trust Scheme Stability

The trust scheme stability suggests that if an entity having stable properties for a significant amount of time (typically until the present time), this proves an evidence for its trustworthiness, since it means that it has reached a state that is good enough to let the entity survive without changes. Stability gives you evidence that this entity status is well-established and reliable; its evolution has reached a mature stage. Instability can be seen as an evidence of an entity that still needs to evolve or correct/change its behaviour and ability. Of course, this is a presumption that must be tested.

Literature supporting the role of stability in trust evaluation is broad. Pickett and Sussman [Pic76] studied the causal attribution between stability and trustworthiness in a complex cognitive framework involving the concept of credibility and objectivity as well. Frewer and Miles in [Fre03] showed how temporal stability is directly linked to perceived trustworthiness. Different bodies – public and private – were asked to release the same information regarding some food hazards. The sample of people involved in the test tended to consider the information more trustworthy if given by a body with temporal stability. For example, hospitals had a higher consideration than the government. This proved that humans consider stability and trust correlated

The Standford Persuasive Labs Guidelines [Sta05] attributes to the permanence and stability of the information on the Web one of the five main sources of credibility and trust.

Computation

In its basic computation, stability can be computing by monitoring how a set of relevant properties P changes in time. Given a set of properties $P=P_1,\ P_2,...,\ P_n$, that are observable and quantifiable, we could simply study the distribution of $D_p=P_n(t)-\mu(t)$, that results a distribution with null average. We quantify the magnitude of the change by looking at the standard deviation of the function.

This gives a global value about the entity P, while if we are more interested in a local increment we could compute the derivative of $P'_n(t)$, or the function $C(t_1, t_2)$ that for a couple of timestamp t_1 and t_2 gives the difference between $[P(t_1)-P(t_2)]/P(t_1)$ that indicates how much the property has changed in the time interval.

Since stability is important in relation to the present time – when the decision-making process has to happen – we need to put more accents on the recent story of an entity, and wonder if it was stable in recent times.

We therefore use the function $C(t_0,t)$, that gives an estimation of how much the property changed in relation to the present value. Given a threshold T, the value t so that $C(t_0,t) = T$ represent the amount of time in which the entity could be considered stable. The computation resembles a percentile of a distribution function, especially because T represents the percentage of variation between the two timestamps. If we chose T=0,1, for instance, t gives us the time when the properties was 10% different form its present value.

The following are the critical questions attached to the scheme. *Is the entity Active?*

Plausibility increases if the entity is actually active in the environment. In fact, an entity could be stable because it is not taking part in its environment, not interacting or not providing services, a kind of abandoned entity. On the contrary, a stable and active entity is an evidence of its reliability, quality of services, ability to survive in its environment.

As an example, we may think about software that has been released several years ago and it is still very popular and used almost unchanged. Likely, the software stability is an evidence of its

reliability and still utility. On the contrary, old software could be very stable because it has been dismissed.

Does Stability carry on useful information? Is the environment dynamic?

If the environment the entity is interacting in is not dynamic and entities do not change their characteristics or even they cannot change, stability does not carry any useful information. Even in dynamics environments this could happen, even if it is unlikely. A way to answer this critical question is simply to estimate average and standard deviation of the changes among the entity population. The function C can be used for this purpose.

Is ∆T significant?

Again, the time of observation ΔT should be relevant. A test could be the comparison of ΔT with L_x , the lifetime of entity X or better with L_{env} , the average expected time of the entities composing the environment.

Is the entity young or in evolution?

This critical question suggests verifying is the entity under analysis can be considered young and stilling in evolution. If it is the case, its lack of stability is mitigated. On the contrary, changes in an old and mature entity have different value, since they are a sign that something has happened and the new status of the entity should be investigated and evidence provided in order to keep the previous value of trustworthiness. For example, if new software has been recently released, there could be a possibility to find problems and bugs that should decrease in time. The same amount of bugs is a worst evidence for an older software, since we may presume that its bugs were not been solved and they are an indicator of more structural problems.

Is the threshold T a reasonable choice?

The threshold T defines how big the changes should be in the period of observation in order to consider the entity stable. A good estimation could be a percentage of the average value of the property computed over a set of comparable entities, useful to minimize the effect of the confounding variables is neutralized.

Are the properties P relevant?

Finally, as for any trust scheme, property p_n should be relevant. We do not need to know which property P is evidence of the scheme premises and which are not. A relevant property is a property that, if changed, make reasonable to think that the entity can be considered a different one for the trust purpose. We do not care if it is a better entity or not, but only realise that something significant is changed. Not relevant properties may change without affecting the scheme conclusions.

For example, if we are interested in assessing the quality of a piece of information on a wikipage, a change in the text is obviously a pertinent property (information could be now of better or worst quality), while the colour of the font is irrelevant for the trust purpose.

Trust Scheme based on Information Sharing and Social Role

Information Sharing

This class of trust schemes base their presumptions on information that members of the community share. As described in the previous chapter, this is one of the two main computational mechanisms to compute trust, in the form of *recommendation* systems, *indirect experience* or *reputation*. All these concepts are build on information that is transferred between two members of the community (recommendations) or available to all the community (reputation).

In this section we do not re-invent mechanisms well-know in computational trust, but we just redefined them as trust schemes, underlying the source of defeasibility. The discussion performed in the previous chapter about limitations and open issues linked to these two ways of computing trust provide us the basis for defining a list of critical questions.

It is important to underline that computational details about how to compute reputation and recommendation are partially described in chapter IV and goes behind this thesis. We do not propose how to set up a recommendation system. Our assumption is that a reputation or recommendation values are data potentially available to the trustier agent. The trustier may decide to use reputation/recommendations values as a piece of evidence to sustain an argument pro or against the trustee, it the argument results undefeated in a given epistemological state.

Trust Scheme Recommendation or Indirect Experience

The presumption is that we can trust an entity on the basis of what a third-party entity suggests. This presumption is based on multiple assumptions, at least the following two: entities usually tell the truth, other entities have a similar concept of trustworthiness. Recommendation are largely used in trust and proved to be effective in some extent. Recommendations are usually explicit trust value, and therefore they support directly trust (they are a trust ingredient).

A trustier may decide to use a recommendation as an argument to support a trust decision, having in mind its plausible nature that was investigated in chapter IV. The results of that analysis let us define the following critical questions regarding a recommendation value (in the discussion R_a is the entity giving a recommendation to T about B).

 R_i denotes the reasoning-link between trust scheme premises and conclusions. The following critical questions are all undercutters of R_i , the reasoning-link between the premises "a certain level of recommendation" and the conclusion "trust/distrust".

Is the recommendation out of date?

It might be the case that the recommendation given by A about B is obsolete and no longer valid, both because of changing in the environment or in entity B. The question is an undercutting reason of the link between premises and conclusions of the scheme: a high value of recommendation no longer guarantees that the entity is still able to fulfil expectations.

Is the recommendation pertinent to the context where the trustee is interacting?

This question investigates if the *trust purpose* of the recommendation and the one of the situation under analysis are the same, issue analysed by Josang in [Jos02].

Is R_a *trustworthy? Is the* R_a *a good source of recommendation?*

The issue is about the trustworthiness of the source of information and its ability to give recommendation that, as explained again by Josang [Jos98], are two separate concepts.

Do T and B have compatible cognitive models/ideas?

It might be the case that B and T think in different ways leading to invalid recommendation. For example, T gives information that it considers important that B would have considered negligible and so on. The issue is well-studied in the domain of collaborative-filtering applications, which tries to group preferences on the bases of similarities among entities. The presence of such a mechanism can increase the plausibility of the scheme. Another factor to be considered is how much the content of the recommendation is subjective: matters such as movie preferences are clearly more subjective than solve a maths test.

Is the recommendation first-hand information or not?

Information too far from the original source could be affected by strong noise due to differences between the agents composing the chain. This represents an undercutter of a computation's assumption.

Does the recommendation received by multiple sources agree?

The existence of a consensus might increase the plausibility of the recommendations; its absence increases the uncertainty. The question represents a potential assumption —or supporter—of the trust scheme computation.

Is there any reasons why entities should provide good recommendation and not malicious?

In the environment there could be a reward for good recommender. A good reason to provide correct information could be represented by the reputation of the recommender entity. Entities with good reputation – especially when this is a public value - have extra motivation to keep their good values and thus they should be more likely to provide correct information about entities in the environments.

Some environments may have random controls over the recommendation shared, or maybe there are punishments for entities sending malicious information. All this extra information can affect the plausibility of the transmitted value. We note how in the digital world some of this information translates into understanding if the system has a kind of central or control authority or it is an open and decentralized one. All these arguments are undercutters of the argument expressed in the critical question, that is an undercutter of R_i .

Is there a positive bias?

As Dellacrois [Del03] noticed for eBay feedback system (see chapter IV), information shared could show a positive bias. The trustier should check if a sample set of recommendation he knows suffer from this kind of bias or the recommendation values distribution are equally spread. This phenomenon is more sensible with reputation, i.e. with a public value, and it has been encountered also in our evaluation described in the next chapter. In the online community analyzed, the reputation values of the great majority of the community were high or excellent, while it was difficult to find members with low reputation. Again, the question undercuts R_i Is there a link between agent A and B?

It may be the case that A and B has a special relationship, being this friendship or conflicts that may introduces a bias in the value of the recommendation. A good test could be to know if the two entities interacts frequently, and the outcome of such interactions.

The above critical underlies the presumptive nature of recommendation, that is too often applied without any consideration of its validity. All these observations are linked to the problem of the small world hypothesis that is central to the plausibility of social-based trust scheme.

Testing the small world hypothesis

The small world property was described in chapter IV section 4.1.4. Here we want to define a method to verify this hypothesis, when enough data are available.

In order to verify the small world hypothesis, we need to verify that

- 1. The length of a random path between two entities increases with a logarithmic complexity
- 2. The community has a low degree of connectivity

Note how the first hypothesis is needed to be sure that the transitive chain connecting two entities reasonable short, avoiding transmitted value to be too noisy to be considered valid. The second hypothesis is not necessary to have a small-world network, but, if verified together with hypothesis one, implies that entities are distributed in clusters connected each other, and few entities are acting as links among different cluster, resulting central to the network, and peripheral

entities only locally connected. On the contrary, if all the entities are well connected, there are no clear division between central or peripheral entities.

A way to computationally verify hypothesis 1 (when computationally feasible) is to collect a sample of random couples, and computes the average length of the path connecting the two entities. Another quicker approximation could be collecting random entities from a population of size N. Then, by starting from each entity A, compute the number of distinct entity reached starting from the A in log(N) hops, and comparing this number with N. If numbers are comparable, the hypothesis 1 is verified.

A way to computationally verify hypothesis 2 available is again to consider the size of the community N. A maximum of N²-N connections are possible when all the entities know each others (note that is the interaction is not symmetric). Then, we count the couple C of entities for which there is at least a connection and we compute $R = \frac{c}{N^2 - N}$, which gives the proportion of couple of known entities among the possible. A low value of R, and hypothesis 1 verified, suggests a population divided in isolated groups or clusters loosely connected, and makes more effective the identification of those entities that are the glue of the community, that connects clusters and are connected to all the entities inside a cluster.

In the case of authority, a small world makes the presence of highly linked entities more important.

Considering the distribution of interactions that each entity did with others may also help the analysis, since the standard deviation may reveal if an entity is connected to few or many others

Trust Scheme Reputation

Reputation, as defined by Sabater [Sab01] [Pin07] is the voice of a community that express its common opinion regarding an entity. Reputation is a more compact value than a collection of recommendations; it is public and therefore more transparent and accessible. According to this trust scheme, entities are trusted for their reputation value

Many critical tests identified for recommendation can be used for reputation, such as the agreement on the reputation values, the understanding of the difference among agents composing the community, the context the reputation refers to. Other critical questions identified from literature are now listed, remembering how the list does not claim to be comprehensive but enough to study the defeasibility nature of the scheme.

Is the reputation system accepted?

It is necessary to check if the existing reputation system is accepted by the community. It might be the case that it has no credibility and its values are completely ignored by community members. An example is described in our evaluation chapter.

Is there a positive bias?

The same reasoning we performed for recommendation is for reputation even stronger, since the fact that reputation is a more visible value; entities are more likely to find reciprocal agreement not to hurt anybody. Since reputation is a public value, entities are likely to do effort to improve their reputation and they will not allow a public display of a bad value.

It has been showed how the fear of "possible revenge" affects the eBay feedback systems. As noticed by Dellacrois [Del03].

Are name-changes easier and cost-free?

Friedman and Resnick noticed how, in the physical world, name-changes have always been possible as a way to erase one's reputation. The Internet highlights the issue, by making

name changes almost cost-free. This creates a situation where positive reputations are valuable, but negative reputations do not stick. [Fri99].

Another side-effect that can defeat the scheme is reciprocal votes among the entities. Koller, one of the founders of the website epinions.com [Epi00] described the situation as follows: "There is a lot of 'You scratch my back, I'll scratch yours,' and mutual admiration societies. You recommend me and mine; I'll do the same for you." A way to test this possible side effect is again to check the distribution of reputation to see if average reputation values are suspiciously high. The information described in the next critical question may also help defeating this undercutter of R_i .

How is the reputation computed?

Information should be collected about how reputation is computed. As any trust scheme, the computation has assumptions and data on which it is defined. Useful information to assess plausibility is:

- if votes have different weight according to the reputation of the voters, in order to selfenforce the reputation value and limit sibyl attacks to entities reputation
- if multiple votes to avoid that few people increase the reputation of an entity to high degree,
- the number of votes and the people voting for the person should be known, in order to investigate the reciprocal vote effect

Trust Schemes linked to Social Role

Trust Schemes linked to Social role base their presumptions on the connections that the trustee entity has in the environment. The scheme suggests that a trustee should be judged not in isolation but for the links and roles he has in the environment he is interacting in. Others may guarantee for him, or its public role/information may give assurance that the entity is not unstable.

The core information to be collected is: trustee's acquaintance, to whom it is linked and interacts, if it has specific roles in the environment, how easy is to access and contact the entity and how transparent is the information he provided.

Among the models reviewed in chapter IV, the sociogram of Sabater [Sab01], the social network-bases application and some trust factor by Carter strongly informs the definition of the following schemes.

We note how the information contained in these schemes is essentially important ingredients to form a value of reputation. As Carter wrote 'the reputation of an agent is based on the degree of fulfilment of roles ascribed to it by the society'.

These schemes investigate evidence linked to reputation that might not be available as a single value provided by a system already embedded in the domain. Therefore the schemes are an alternative investigation that can confirm or contradict the results of previous schemes in accordance with the argumentative and defeasible nature of our model. Again, the mutual relationship between the schemes will produce a final clearer value to support decisions.

Trust Scheme Authority

According to this scheme, that is a fundamental ingredient of reputation and therefore of trust, an entity is trusted according to the importance that other entities assign him (or to its actions). Authority is usually based on a strong consensus, few counter arguments or alternatives, and high popularity. Authority is also a classical example of non-monotonic scheme. Authority is

linked usually to the truth of some assertions that, in the light of new evidence, can decrease or completely revert the authority of an entity.

We separate this trust scheme from reputation for its huge success – and abuse – in computational trust. Authority of entity A is usually computed by collecting implicit evidence about the recognition that other entities assign to A.

The presumption underlying the argument is that the presence of these pieces of evidence means that entity B trusts A, or it recognises the authority of A, its ability or importance.

The most common implementation of the authority scheme is the citations-based one, or the analysis of the linking structure among the entities.

In practice, pieces of evidence used are citations, references, link to an object related to the trustee entity or itself. Great part of the plausibility of the scheme is based on the presence of these pointers from entity A to B, and the presumption that their presence act as an implicit recommendation and recognition of B's quality by A. Usually this pointers are used without critically testing their meaning.

Computation

The computation performed is usually based on the analysis of the quantity and distribution of the pointers from A to B. Authority of A increases if A has been linked many times, by many different users. It is better if A is linked both occasionally and repeatedly. It is important to understand if the distribution of pointers is not biased by a restricted group of users and, on the contrary, it is not occasional.

Therefore, a possible computation of the trust scheme for entity A may consider the value of C(n%) of the ranked distribution of entities ranked by the number of pointers to A, or the comparison of average and standard deviation, to understand the distribution of the pointers by users. C(n%) represents the number of pointers of the first n% of the entities linking A. The use of C(n%) assumes that we discard contributions outside the top n% entity in the rank. This assumption may be wrong when there are enough evidence to consider even the entity that interacts with A occasionally.

Critical Questions

Which is the meaning of the act of linking A by B in the domain?

This is the essential question. It requires investigating the relation between the action of *linking* and its meaning as a good/bad implicit recommendation. How many times the presence of pointers indicates judgements that can be linked to the notion of trust?

The defeasibility of this hypothesis undercuts the entire scheme, since the reason linking evidence to conclusion is invalid. For example, the *PageRank* algorithm is one of the most successful and used. The presumption is that if website A links website B, A's webmaster has recognized that B is somehow useful. It could be because it is an interesting website, or because is the leader on the topic A is about. B could complete the content of A, but among the set of possible websites to be linked the webmaster choose B.

These observation increases the plausibility while others are decreasing it. For instance, the fact that there is no alternatives for linking a website, or A is obliged to link B, or it does because A and B are somehow linked, or just to increase artificially the authority of a link.

When the plausibility is low, a trust computation based on the presence of these pointers is clearly useless. Our evaluation chapter shows how *Pagerank* could have no meaning in specific context, even when entities are highly linked and interconnected.

How frequent is the action of linking? Is it supposed to be repeatable?

If the action of linking unique and not repeatable, the computation should consider this information. If the action of linking can be repeated, we have to wonder if a repeated link means stronger evidence, in bad or in good. The question represents an assumption of the computation. For example, if an author A quotes many times another one, and in different situations, this is strong evidence that A recognises the authority of B. In this case, in the computation the use of the indicator C(n%) helps to weight more the entity that gave multi-contributions *How big is the entity space? How entities are connected?*

If the entity space is big enough, it could be more difficult to artificially manipulate authority, since the environment is harder to be controlled.

These could avoid the situations where the number of links required to gain a good value of authority is quite low and therefore manipulable in a way similar to how reputation can be artificially build. Another important factor is the way entities are distributed in the environment: do they form cluster loosely connected to each other or do entities have a uniformly connection with all the other?

According to this feature, the value of authority could be increased or decreased: authority recognized by members of different clusters is more effective than an authority localized on a single cluster.

A fast way to quantify it is to compare the size of a cluster the entity is in (or the average size of the cluster) and the number of pointers. Again, the question underlines assumptions of the computation.

Confounding Variable analysis

Confounding variable could be the age of the entity and the context/topic for which it is linked. A young entity may require time to gain authority, or an old entity may have a higher computed value of authority that is not reflected in actual value of trustworthiness.

Particular attention in the computation should be given to not double counting the age of the entity. If the longevity trust scheme is used, authority should not be affected by age and this confounding variable should be isolated with, for example, a cohort experiments.

Trust Scheme Connectivity

The link between two entities could be considered an indicator of a *social* relationship rather than an implicit recognition of authority. We define this variant, more social-oriented, the *connectivity* trust scheme. The presumption is that we trust an entity that shows high connectivity with many other entities in the community. Connections that this entity has, can be simple acquaintance, it does not mean an implicit judgements about entities as in the Authority trust scheme. The presumption underlying the trust scheme is different as well. Here we presume that, because of the fact that a person is highly known in the community, it means that he is more accessible and visible, he has acted well, without betraying anybody, maybe honestly and he fells confident in the community.

The computation is analogous to the trust scheme authority by using the distribution of connections instead of the action of linking.

Critical questions used for the trust scheme authority concerning the repeatability of the interactions are applicable in this context as well. The repeatability can strengthen the relationship among two entities – there is not an occasional acquaintance but a friendship.

It must also be checked the plausibility of the elements used to detect an interaction: does the evidence really means that A knows B? And is it possible to discern if their relations is positive

or negative? The repeatability of the interactions and its time-distribution may help; since we could defeasible assume that friends are likely to interact repeatedly, while enemies have less or no reason to do so. Another possibility is to perform a detailed study and classification of each interaction by sampling specific cases to quantify the validity of our presumption.

Trust Scheme Popularity

The trust scheme popularity states that an entity should be trusted because of its popularity, which is a defeasible reason to believe that the entity has good trust ingredients such as reliability and ability to satisfy the needs of the majority of people. It is out of discussion that chart or selling rankings can influence our decision making process when we are close to buy a new item, such a car or even a music CD. Relying on the power of the majority could be a dangerous presumption: information should be collected about the subjectivity of the community taste, the motivation for the entity to be so popular, the alternative possible — an entity could be so popular because there is no competitors - and investigate the effect of confounding variable such as time, price. All this critical observations represents undercutters of the defeasible reason encoded in the trust scheme.

Popularity is a trust scheme clearly connected to social role and recommendation, but it is also connected to some extend to the degree of activity of an entity, and it is therefore analyzed also later in this chapter when we speak about the activity trust scheme.

Trust Scheme visibility/accessibility

According to this trust scheme, an entity increases its trustworthiness by thanks to visibility and accessibility, clearly linked to the trust ingredient accountability. This means that an entity reveals personal information that makes him more transparent and easy to access for the other member of the community, this, as stated by Romano [Rom03], has been proved to increase the perceived reliability and trustworthiness of an entity. For example, in [Eco06] it is proven how poor accessibility of web sites can strongly affect customer trust, as a result of experimentation over 500 UK website. Lack of visibility or accessibility decreases the transparency of an entity and call for extra investigations.

The quantification of accessibility/visibility usually requires testing the presence of information linked to entity's identity and contacts, or when the entity was accessible. For example, in the context of an online community, we can compute the percentage of time a member of the forum was online, or we could evaluate the presence of link to personal information on his profile page. Even if the trust scheme could be very effective as underlined by social scientist — one of the main component of trust is transparency that increases the perceived sense of control to use the words of Romano [Rom03] — it is usually hard to perform a study of the validity of the information provided, since in many cases it is impossible to verify it and therefore assigns a value for the trust scheme. This represents a direct defeater of the scheme.

When it is possible to access the required data, a possibility is again sampling and investigating the validity of a subset of the data available in order to have an idea of its validity, as we perform in our evaluation. The percentage of valid information given by the sampling will determine the defeasibility of the scheme: a percentage of 50% carries no information, while a high percentage close to 0% or 100% reduces the uncertainty.

Trust Scheme Transitivity

Among the most used in computational trust, the transitivity trust scheme presumes that trust can be transferred among trusted entity: if A trusts B and B trust C therefore A trusts C. Transitivity is *per se* a form or implicit recommendation, where the trustier exploits other entities information in order to compute a trust value for the target entity.

Transitivity has been largely discussed in the previous chapter. Example of computations have been presented – (Golbeck, Poblano, Sierra), here we remind its plausibility nature listing some critical tests. Note how these tests are directly linked from open issues discussed in chapter IV.

Critical test for recommendation involving the subjectivity of the entities involved in the computation are still valid here. Moreover we have to include the following (all computation assumptions to be made explicit to defeat/support the scheme):

How long is the transitivity chain?

The longest is the transitivity chain, the noisiest are the results. Many authors, such as Poblano and Sierra, encompasses this information by inserting in the computation the length of the chain that has the effect of reducing the plausibility of the transmitted value (see section 4.1.3).

Are there cycles?

The presence of cycles can artificially alter the results. The problem has been already treated in chapter IV.

Is there evidence of referral trust ability of the elements composing the chain?

The entities composing the transitivity chain has to be good in referring valuable information about other trustworthiness, not being a trustworthy entity by itself. The way of testing this issue is described in Chapter 4, and it is common to recommendations as well. *Is the transitive chain valid?*

The concept of valid transitive chain has been defined by Josang (see section 4.1.3). A valid transitive trust path requires that the *last edge in the path represents functional trust and that all other edges in the path represent referral trust, where the functional and the referral trust edges all have the same trust purpose.*

Is the small world hypothesis tested?

Analogous to other social-based trust scheme

Is the transitivity trust-based or social-based?

It should be understood if the type of information shared in the transitivity chain is explicitly referred to a trust purpose, it is a generic trust value or it is a social-based type of information. This distinguishes Social Network from a transitive chain or recommendations. If the information shared are based on a social-relationship, such as A is friend of B, we should wonder if from this type of information – usually a single generic value - we can derive trust information

The questions undercutting the scheme are therefore: *could we switch from acquaintance to trust? Do they relate to each other?*

This could be true in many cases, but it is easy to show that in general they are not necessary conditions in both ways and therefore it represents an undercutter of the scheme.

Trust Scheme Information Provision

This trust scheme is linked to one of Carter's trust factor [Car02]. The description given by Carter is the following, already discussed in chapter IV: 'Users of the society should regularly contribute new knowledge about their friends to the society'. This role exemplifies the degree of connectivity of an agent with its community. The degree to which the social information provider

role is satisfied by a given user is calculated as the summation of all its recommendations/piece of information mapped in the interval [0,1].

The quantification of this trust scheme is therefore based on the quantification of the useful information provided for the rest of the community. For example, in an online forum this could be mapped to the amount of supporting information helpful for another community member.

Other indicators of this trust schemes are the number of time and frequency each user votes/recommends/gives feedback about another one/entity in the environment, since it is accomplishing its social task of providing information about other entities for the benefit of the community. Note that it is not necessary to know the content of the recommendations or feedbacks.

The same arguments discussed for the trust scheme visibility are still valid for this trust scheme, in particular the difficulties of performing an (automatic) analysis of the plausibility of information that is not strongly time-consuming and maybe not feasible.

Trust Schemes linked to Activity

This group of trust schemes focuses on the activity of entities in the environment, i.e. what entities did rather than when or how. It focuses mainly on quantitative aspects, not considering outcome of an action. Information such as type of action performed, its context and topic are also considered, but not the outcome or the quality perceived by other entities. Rather, they are based on a classification of the type of actions, an investigation of their complexity and pertinence to the context.

In the trust scheme pluralism, we focus also on the comparison of amount of activity of different users in the same context.

The basic presumptions are that an entity cannot be trusted if it is not active in the domain – this casts a doubts that should be investigated—while the high activity of an entity in a domain could be regarded, under some conditions, as an evidence of a good health status, high reliability and success in interactions. These set of trust schemes are an alternative to the classical duo recommendation/past-outcomes.

Trust Scheme Pluralism

The presumption could be expressed as follows: "I trust what is the result of many entities cooperation". The grounding pattern of the scheme is the following. An element X (being an entity, an action, a property) is affected by an action A done by a group of entities Y_n . We presume that the resulting element X is trustworthy because it is the result of a collaborative activity of multiple entities and points of view. Pluralism should guarantee X to be less biased and more objective, representing direct reason to trust. The scheme applies naturally in content-quality assessment.

Note that, in order to identify the pluralism pattern, we do not need to know the quality of each entity's contribution; we do not need a judgment over its actions. This is a core feature of pluralism; if we had to collect a set of judgments, we would have built a recommendation system. The scheme presumes that element X is the result of many entities cooperation and this evidence alone, despite the different quality of the single contribution, guarantees that the resulting entity, as a whole, is more objective and less biased. Of course, some conditions make this assumption stronger or weaker.

Computation

In order to be applicable, the trust scheme needs a situation in which the action by a set of entities Y_n determines the status of an entity/object X. The action performed by Y_n must be observable and, more important, it must be quantifiable, so that it is possible to build at least a partial order on the various contributions of Y_n .

Given these conditions, the scheme can be computed by observing the distribution of the contributions for each member of the set Y_n .

In quantifying pluralism the number and size of size of each contribution should be considered in the computation's assumptions.

A better pluralism (still defeasible) is obtained when contributions are not coming from a single or a very limited number of individuals — leading to a dictatorship effect that can affect the neutrality of the resulting X — and on the contrary contributions are not a plethora of small insignificant ones, leading to a fragmentation that can decrease the consistency of X, its cohesion and therefore its reliability. At least some leading entity should be identified.

If we rank the distributions done by Y_n , and we call this distribution $D_r(Y_n)$ pluralism can be computed by the usual statistical indicators such as:

- Average and Standard Deviation, the latter revealing if there are members of Y_n that contributed much more than others
- C(n%) or C(n): that give a quantification about how much of X is the product of the top n-top entities in respect to their contribution to X
- 1-C(n%) or C(N): it quantifies the contribution of entities with low participation

In order to avoid fragmentation, the above indicators should not be too small, except for the last one that should not be too high. The opposite stands for the dictatorship effect.

Pluralism is a computational mechanism that goes across different schemes and actually it acts as a critical test (an *undercutter* argument) for many of them. Note how pluralism is actually a computational mechanisms used by several trust schemes. If action A is the act of giving recommendations, the pluralism increases the recommendations result. A global reputation value is more plausible if its pluralism is higher. *PageRank* is also based on pluralism.

Here we intend pluralism as linked to actions that create objects rather than judgements, where each contribution is visible. Due to its importance, we gave it the status of distinct trust scheme, pluralism is a core aspect of any generic collaborative environment, encompassing a large class of applications such as self-organizing system of information like wikis and forums.

We now list the set of critical questions leading to undercutter arguments: Are the entities Y_n recognizable?

Entities should be recognizable, otherwise we cannot distinguish them and conclusion based on can be diminished or even undercut.

How is the cardinality of the set Y_n ?

The computation is more sustainable when the cardinality n of the set Y_n is reasonably high in relation to the problem space and the number m of entities whose contribution to X was not negligible is significant. Note the statistical indicator C will quantify this What's the frequency F_a of the action A?

Action A should happen with a frequency F_a that guarantees that many outcomes can be collected in the time interval of observation ΔT or during the lifetime of the entity L_x Are the entities independent?

Plausibility increases if entities Y_n are independent (think about the same information confirmed by different independent source) or there is not an explicit relationship among them *Is the action A critical for X?*

There should be a clear link between how *A's outcomes* influence the status and therefore the trustworthiness of X.

Is the quantification of A plausible?

This test focuses on the way A is quantified, on the criteria used to order two dfferent actions. The criteria are a computation's assumptions that affect its validity.

Finally, further tests require analyzing how the process that produced entity X was carried out by entities Y_n , i.e. the dynamic of the process. The more the process is dynamic, the more information about pluralism can be collected, while if the process defined by action A has some constraints, pluralism plausibility can be strongly affected. Among the test to be considered:

Are contributions occasional or can they be repeated?

Can an entity collaborate more than once with X?

If it cannot, every argument related to fragmentation should be revisited, since under this condition fragmentation cannot be avoided and the trustier should question the meaning of Pluralism in such a context.

Is it possible to cancel or revert the contributions?

If it is possible for the entities to make repeated contributions and to revert or cancel others' contribution, more information can be collected on the dynamics that generated X, information that can make the computation of pluralism more transparent and plausible. For example, by considering the number of cancellation and revert actions it could be possible to understand if entity/object X is the result of a linear or discussed process. It is therefore possible to investigate the significance of the dialectical process.

Trust scheme Activity (trust based on the degree of activity of an entity)

This trust factor investigates the use the degree of activity of an entity as an indicator of its good status and therefore a reason to trust. Defeasibly, entities that are not active should be trusted only after further investigations. On the contrary, an active entity represents evidence whose plausibility should be analysed.

A high degree of activity could be an implicit indicator of good outcomes, especially in a high competitive environment. For example, a company that is registering very high selling in a competitive environment could be evidence that the product has good quality, only usng the quantity of units sold as an indicator. The focus is on how much the entity is active, the ype of action performed but not when it acted or the quality of such actions. A high degree of activity entity indirectly increases other trust scheme such as accessibility, visibility, temporal trust factors.

All these presumptions are tested via critical questions. Critical questions analysis should reveal if a high amount of certain types of activity represents plausible connections to trust.

The basic computation of the scheme implies to quantify the amount of activity of an entty in a domain. Usually, the computation ranks several indicators, each of them measuring some spects of the activity. The indicators must be accessible, the action observable and sonehow quantifiable. Uncertainty could affect this scheme more than others. If the quantification is not possible or plausible, the whole scheme will be affected. This preliminary condition, essertial for the scheme to be applicable and it may require some domain specific expertise to define the types of actions and a possible ordering function for each type. The followings are the critical question identified:

Which is the relationship among the evidence found? Is there more than one type of action?

Usually the final value of activity is an aggregation of several indicators. At entity activity in a domain could be represented by more than one type of actions, such as in ar online

forum an entity can post a message, open a thread, attach images, or a user during a browsing session could open a page, *bookmarking*, submit a query, *cut&paste* text.

The way we aggregate these elements into a final value is not trivial, requires domain-specific expertise. The problem is to define an aggregation function F_{agg} for the premises of the trust scheme.

The rankings can aggregated by using one of the strategies described above, on the basis of an analysis of the type of actions identified and its features. For instance, action A and B can be either necessary, or B can be an optional action that can increase the quantity but not decrease, or A or B are exclusive. For example, as shown in our evaluation chapter, in the context of an online forum the action of posting a message is considered the essential base-action, while actions such as inserting an attachment can clearly increase the amount of activity but it is considered a plusentities are not obliged to provide constantly attachments - while entities that do not post messages cannot be considered active. In all our evaluation, the F_{agg} used are one of the type defined in the previous chapter and we do not use ad-hoc functions. A good or bad aggregation represents a potential undercutter or supporter of the computation.

The following critical questions are related to a kind of Market-like dynamics that might be present in the domain, and they all represent undercutter links to the reasoning-link between premises and conclusions.

Is the environment competitive? Is the entity in a privileged position?

If the environment is high competitive, and entities could have different choices, a high degree of activity is more plausible evidence. If a bakery shop is the only one in town, it is in a privileged position to sell its items. On the contrary, if the shop does not sell enough units – people prefer to travel to other towns to buy bread – this is a plausible indication that the service offered is poor.

Are entities actions in competition?

This question is again about competitions. We wonder if an activity performed by an entity X will result in less activity for another entity. In other words, we wonder if the total amount of activity is somehow fixed an entity has to compete for them, exactly like in a free market dynamic. The question generates an argument that might be a strong supporter or defeater of the reasoning-link between premises and conclusion of the scheme. High activity in high competitive market is a strong evidence of entity's reliability and ability, while high activity in not competitive market is a weaker argument, but still not negative as a lack of activity would be. How is the size of the market?

In order to finish this set of market-related questions, we wonder if in a competitive environment the offer for a service is so high that all the *seller* entities are active independently from their ability.

How is the market trend?

An analysis of the degree of activity over time can reveal the trend of the market, if it is in a contraction period or expansion, argument used to support/defeat a given value of activity. Which is the size of the entity?

An entity may have limited resources, so that it cannot be active as other bigger entities. Nevertheless, it may show a high degree of activity in relation to its size. Size or quantity of resources is therefore a confounding variable to be taken into account. For instance, in the context of the Wikipedia project, we can't expect that an article covering a popular and important topic such as *USA* can register the same number of access and editing than a nice article about a minor geographical location such as the town of *Salò* in the North East of Italy. The same reasoning

stands about the age of the entity. The key-point is that age and size are two potential confounding variables that should be isolated.

Does the entity need to be active to survive?

If an entity that is not active simply disappear from the environment - exactly like a shop that does not sell – the degree of activity is an indicator that the entity is able to survive, (defeasibly) because it knows how to perform well in the environment

Is activity supposed to be continuous?

As we perform with the time-based trust scheme such as Persistency, we wonder if, during the time of observation ΔT , activity is supposed to be continuous, or it is cyclic. If it is the case, ΔT should be large enough to cover cycles in order to avoid collect level of activity too high or too low because of the presence of cycles. Therefore, time could help to avoid another confounding variable.

IS ΔT enough to quantify an entity activity?

The lifetime L_x of entities and the frequency F_a of the actions can test this hypothesis. If $\frac{1}{F_a} \ll \Delta T$ and ΔT is comparable to L_x conclusions should be plausible, otherwise ΔT may not be enough to understand the real amount of activity if entity X.

An essential argument related to the defeasibility of activity is the analysis of the pertinence of such activity. Therefore the following critical question is crucial:

Which is the pertinence of the type of actions? Is the activity pertinent to the trust purposes of the context/community? Is the activity competent?

The analysis of pertinence is an essential argument to support activity as a trust scheme. We need to understand if the action performed is pertinent to the trust purpose of the trustier.

It may be the case that an entity has high activity in a trust purpose that cannot be linked to the one the trustier is interested in, and therefore the amount of activity is invalid evidence. The more the information is pertinent, the more the entity is presumably competent in that specific trust purposes. We remind how competence is regarded by many authors as an importance ingredient of trust (see [Cas00], [Lag96]).

For example, in an online forum about trading online, an entity that constantly writes messages about non-trading topics is less pertinent than a member posting news, graph analysis, and its amount of activity, even very high, it easily defeated as trust evidence.

A pertinence analysis requires quantifying the amount of activity that is pertinent to the purpose, it implies usually to classify which type of activities are considered pertinent, or define indicator of pertinence to be applied to the set of entity's actions.

Due to its importance, Pertinence could be considered also separate trust scheme that is linked to the trust scheme activity by an undercutting link that can defeat or support the value of activity.

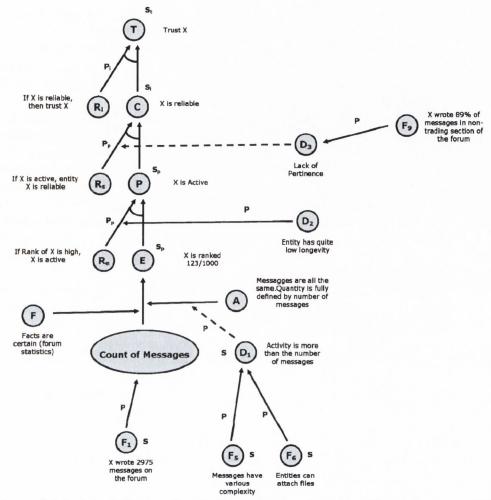


Figure C.2 Activity Inference Graph. Example taken from Experiment I, chapter VIII

Trust Scheme linked to outcomes

The most used computational trust mechanism, the past experience scheme presumes that an entity that acted well in the past, producing the expected outcomes, will continue to do so. Therefore, an entity is trusted for its past achievements as sign of its present ability to fulfil. Form a computational point of view, in the previous chapter we described various method for computing a trust values based on past experience that resembles a learning strategy enforced by feedback loop.

The scheme is regarded by some authors – such as the $\mathit{Trustcomp}$ group [Tru04] - as an objective measurement of trust. Here we show its deeply defeasible nature that can sometime invalidate completely the scheme.

We assume that the outcomes are objective; therefore it is possible to discern if an outcome is good or bad. We avoid the problem of subjectivity since it has already been treated in the context of recommendations and transitivity. Since trust values computed by past-outcomes are usually transmitted to others via recommendations, or added up in a global reputation value, the problem is clearly central but contained in those other trust schemes.

We also removed the uncertainty that derives from the fuzziness that some outcomes may have, so that even the trustier has not a clear idea about its own satisfaction.

Here we assume that the past outcomes are collected by direct experience by the trustier entity, which knows how to evaluate them, removing the fuzziness that may arise from this other issue. Even these hypotheses leave the scheme highly plausible. Before starting our discussion, we define some terminology.

 T_r is the trustier entity while T_e the trustee that in the past performed n times an action A with outcomes O_n . S(O) is the function of satisfaction of the trustier that, given an outcome, returns its level of satisfaction. The action A have a frequency f_a . The period of observation is ΔT

The past-outcomes trust scheme is applicable when the outcomes O_n are observable by the trustier and the function S exists, i.e. outcomes can be linked to degree of satisfaction and a computation such as the ones described in chapter IV may be used.

But which critical questions may affect the plausibility of the computed trust value?

In a paper called "How trust is influenced by Past Experience and Trust Itself" [Fal04] Castelfranchi and Falcone show complex factors behind the direct experience trust scheme, that in its basic form they consider naïve. They wrote:

It is commonly accepted that one of the main sources of trust is the direct experience. Generally, in this kind of experiences to each success of the trustee corresponds an increment in the amount of the trustier's trust towards it, and vice versa [...] There are several ways in which this qualitative model could be implemented in a representative dynamic function, [..] but this view is very naïve, neither very explicative for humans and organizations, nor useful for artificial systems, since it is unable to adaptively discriminate cases and reasons of failure and success.

By adding our contributions to their discussion, we define the following list of critical tests partially already described in chapter III.

Is the information out of date?

If the past experiences collected are too old, they could be invalid. If entity E_1 did very well in 3 occasion in the past, but performed badly recently, and entity E_2 did the opposite, are we sure that T_v of E_1 is greater than T_v of E_2 ?

The problem is the selection of the proper ΔT , period of time in which the past experience are kept. This could be embedded into the assumption of a computation that consider the age of the past experience, or it can represent an undercutter of a computation that does not consider evidence age.

In the choice of ΔT , it is too big, trust value is slow to react, and an entity that did well in the past can rely on a high trust value even if it performed bad recently, while a too short memory does not take in consideration past experience and it is very sensible to occasional good or positive outcomes.

A reasonable ΔT could be selected by looking at the frequency f_a of the action A and the lifetime L_x as already performed for other trust schemes.

Is the action A or the environment changed?

Action A to which the outcomes are connected may be different, or the environment changed. For instance, Bob is a car mechanic that proved to be very good in repairing mechanics-based car. With the introduction of electronic-based cars, the task is now completely different.

Of course part of the trust can be "kept" in the new context, thanks to the ability that Bob showed, but it is out of doubt that evidence is weaker than before.

Has trustee T_e changed?

Trustee's properties may have changed, so that past experiences are invalid. Stability should be check under the period of observation. Again, this is a cross-relation among trust scheme.

Is the number of collected outcomes enough?

The more outcomes we collect, the less is the uncertainty associated with the final trust value, as many authors such as Quercia and Josang already inserted in their computational models. This could represent a strong supporter or attacker of the scheme. Again the analysis of f_a , the lifetime L_x of an entity and the period of observation ΔT has to be considered.

There is a set of critical question that investigates the reasons for a bad or good outcome others than the ability and competence of the trustee, reasons that are out of the trustee's control. On this topic, the authors wrote as:

We challenge the trivial idea that always success increases trust while failure decreases it. Of course, this primitive view cannot be avoided till Trust is modelled just as a simple index, a dimension, a number; for example reduced to mere subjective probability. We claim that a cognitive attribution process is needed in order to update trust on the basis of interpretation of the outcome of A's reliance on B and of B's performance (failure or success). [Fal04]

A cognitive attribution, a more depth interpretation that in our method is represented by the critical questions, is needed before using computed trust values. The critical questions to consider are:

Was the outcome of action A affected by external constraints outside trustee controls?

It may be the case that the failure of the trustee is connected to external environments or constraints that are not under the control of the trustee. External constraints a successful outcome count less than a success in hard times and vice-versa, failures in hard times are mitigated, failure in favourable times are highly negative.

An example taken again from our evaluation chapter refers to stock market predictions: it is easy to have success in years of bullish market and harder in time of bearish market, even if the entity under analysis is a very skilled trader. Castelfranchi and Falcone suggests also to consider if these factors are external or internal (due to something happened to the trustee) and stable or occasional.

Does T_e have the motivation to accomplish the task?

This is another important cognitive mechanism, cross-related to the trust scheme motivation described later. If the trustee does not have the motivation to accomplish the outcome – maybe because it wants to hurt the trustier or because it has more profitable interactions to do –, very positive past outcomes are useless. We note how, in any case, the outcome of the current interaction will affect the global trust value for future interactions, limiting future damage.

Trust Scheme based on prejudices and grouping

These set of trust schemes ground their assumptions on the statistical significance of some properties of the trustee compared to other entities or group of entities. The sociological motivation behind these trust schemes are the socio-psychological studies of Kahneman and Tversky [Tve74] while the use of *categorization* in Castelfranchi Falcone [Cas00] and the concept of *prejudice* in computational trust as described in section 4.1.4. Entities trust other entities on the basis of the categories they belong to (or they are supposed), or on the basis of similarities/dissimilarities with the trustier entity. The common idea behind these mechanisms is that trust can be transferred among similar entities/situations. Therefore, entities are pre-judged for belonging to a group. We remind how in digital world prejudice does not have a negative

meaning but is the mechanism of assigning properties to an individual base on signs that identify the individual as a member of a given group [Sab05].

These groups of trust schemes encompass *Similarity*, *Similarity to Trust*, *Categorization* and the *Standard compliance* trust scheme. They are all base on the same concept of similarity quantification, but the first analyses the similarity between the trustee and the trustier – therefore reflects a local point of view -, the second the similarity between the trustee and the *stereotype* of the trustworthy entity that the trustier build in its mind. The third scheme assess the similarity between the trustee and a group of entities and the fourth the similarity between the trustee and a standard –if any – that emerged or is defined in the entities community or in the mind of the trustier. The way of computing them and the critical tests associated are similar, so we describe the trust schemes all together.

Similarity, Categorization, Standard

The standard compliance trust scheme has a statistical nature with a clear human related meaning. The assumption is that entities showing properties significantly different from the average of their categories or from a defined standard should deserve further investigation and it is not prudent to grant them trust without further evidence. If a system is not compliant to a standard, that could be an indicator that it is the product of a process that may not be aware of the state-of-the-art, or produced by non-experts or in an unchecked or asystematic way. In the context of computational trust, analogous mechanisms were used by Ziegler and Golbeck [Gd00] while Castelfranchi and Falcone underlined the importance of categorization as a first basic form or reasoning in trust [Cas00].

Regarding the trust scheme similarity, it states that because of the similarity of two entities a rust relationship could exists, since the entities have a special inclination towards similar entities. *Computation*

The way these trust schemes can be quantified is straightforward, due to their statistical nature. *Similarity* can be computed with a measure of distance in a multi-dimensional space between two entities' properties.

The *compliance* to a standard or *categorization* of an entity can be quantified by looking a the basic statistical indicators of the distribution of entities' property, such as the distance forn the average in terms of standard deviations. If an entity has some properties that make him an oulier, it should be investigated why he is in that position.

This trust schemes show all its statistical nature, resembling the way control charts are used to monitor the performance of a system and identifies the outliers. When a distribution can be threaten as normal, the possibility that a measure is far from the average value of a quantity equals to 2σ is 5% and 3σ is 1%. This information is used to monitor possible failure or anomalies in the systems.

In Walton and Cobb [Wal96] this scheme finds correspondence in the argument from popularity and in the argument from analogy. Kahneman and Tversky [Tve74] largely studied how similarity and categorization are cognitive mechanism used to make judgments under uncertanty. *Critical Ouestions*

Choice of properties

A crucial point of these schemes is the choice of the properties on which the computation is performed. In the case of the standard compliance mechanisms, the presence of a standard suggests the features to be considered.

Regarding similarity and categorization, what is important is that the properties chosen give an appropriate and exhaustive description of the entity in its domain. The assumption of the trust scheme, proven plausible both by social scientist and computer scientist, is that it is possible to correlate trust to similarity, and therefore to deduct from a relation of similarity a relation of trust. This is exactly the presumption of the schemes. The choice of the correct properties can be discussed in the same way we did for Stability: the properties chosen have to be significant in determining the status of the entity, not directly linked to trust.

It is certainly important to select a proper range of similarity indicators, that may involve things such as interests, making a link to collaborative filtering methods, but in general elements used to compute similarity are not judgements-based, but they are linked to descriptive properties of an entity visible in the domain, linked to the core-activities of the domain.

As an example, we apply the scheme in the context of the Wikipedia articles. Properties such as text, images, presence of reference or sections have been used to describe an article and apply the similarity trust scheme.

Is there a standard defined?

If in the domain a standard has been defined – by law, consensus, emerging guidelines – and all the entities should resembles these standard, the standard compliance trust scheme's plausibility increases.

Why an entity is not compliant to a standard?

It should be considered if entities are allowed to be different from the standard. It should be taken into account the age of the entity and its degree of activity in order to verify potential reasons for the lack of standard.

Is the difference in positive or negative?

It is important to know if entities properties are better than the standard or worst. This analysis implies a more detailed investigation of the domain, and collecting domain-specific knowledge in order to understand how to order entity's properties. Note how we do not need to understand that, whether A>B then trust of A > trust of B, the link with trust is not a problem. What needs to be understood is whether A>B implies that A is fulfilling the defined standard more than B.

It could be the case that the standard is defined by minimum values that entity should exhibit, solving the problem of properties ordering.

Are the properties observable and credible?

It should be taken in consideration if the properties we use to test similarities are transparent or if there is an uncertainty associated with them

Size of the population

The plausibility of the schemes increase if the set of individuals is numerous, since the central values become more solid, less uncertain and the outliers are more evident.

Is the group compact?

A compact group, with low variance, makes the presence of outliers more suspicious and the standard more plausible, while large variance may imply the non-existence of a standard or a defined category

Is the group evolving?

The plausibility of the scheme increases if the categories we are using are well defined, the entities have comparable lifetime in the environment and the set of properties p is stable, meaning that it is not in the process of evolving and thus with unstable average values.

Similarity to Trust

Particular attention is given to the similarity-to-trust scheme. According to this scheme, an entity is trusted if it is similar to other trustworthy entities. The scheme requires having a notion of trustworthy entities usually maturated using past experiences, and to quantify the similarity of the trustee with this stereotypes trusted entity.

The set of critical questions are not different to the ones presented so far: the choice of the properties that define the trustworthy stereotype and the number of evidence used to define it.

Statistical tools such as the principal variable or automatic features selection can help to identify the set of significant properties that should be used. The scheme requires having a previous knowledge of what is trustworthy to essentially train the system. Therefore, computational mechanisms such as implicit prediction tools may be used, introducing with them all the limitations and assumption that such tools has.

A potential undercutting argument is therefore the objectivity and validity of this previous notion of trust along with the context to which it refers to.

Critical question attached to the past-experience trust scheme, such as the ones related to possible changes in the environment are potential defeater as well, since a change in the environment or entities subject to rapid changes made the defined stereotype out-of-date and no longer appropriate.

Game theoretical/Cognitive Trust Scheme

These set of trust schemes consider opportunistic motivations that the trustier and the trustee may have in the situation, modelled as a game among rational players. The assumptions behind these trust schemes is that the trustee and the trustier are both rational entities that are trying to maximize their satisfaction and minimizing the effort spend.

Therefore, the understanding (or the presumption of knowing) the cost and benefit of the other entities produces an argument in favour or against trust.

These schemes should not be seen as a reduction of trust to a mere quantification of the utility that each party gains in the interaction, since this quantification includes cognitive reasons, such as the motivation of the entity.

Due to the kind of evaluation performed in the next chapter, these trust schemes were not investigated from a computational point of view in this work.

Anyway, we describe them here for giving a list of trust scheme more complete, providing the basic presumption and their critical questions and knowing how our description represents just a starting point.

The exploitation of game-theoretical trust schemes have been already proven by various authors as described in chapter IV. Anyway, we notify more cognitive-oriented schemes – fulfilments and motivation - do not have an efficient computational version so far.

The following schemes seem to be more useful in agent-based situation, more linked to outcomes and interactions than the scenarios used in our evaluation.

Trust Scheme common goal/situation/risk

The presumption is the following: if two entities share the same situation is more likely to help each others. Both the parties have interest in the situation and therefore they may merge their effort.

The following two critical questions are an evidence of the defeasibility of the scheme: *How can we be sure that the other entity is in our situation? Can he prove it?*

In a distributed and argumentative scenario, this could be hard to verify representing a defeater of the argument.

Does the trustee gain any advantage by the situation?

We could wonder if the trustee is actually suggesting something that will modify the situation in its advantage. For instance, two traders may own shares in the same company. A trusts B's suggestions, presuming that B will act honestly because he is involved in the same business

Anyway, B could suggest A to sell its shares only because he wants to buy them, or B may give away false information about the company in order to convince other people to buy it, overestimating maliciously the company. In this case, even if we know without for sure that B holds some shares of the company, this is not a plausible argument to think that B's suggestions can be trusted.

Benefits/Costs Trust Scheme: the trustier motivation

The trust scheme focuses on the classical benefits/cost analysis: a trustee entity should be trusted more if it has a very favourable benefit/cost ratio in the specific situation. Note that here we do not consider the benefit/cost analysis of the trustier, whose discussion is contained in the risk trust scheme explained later in this section. Here the focus is on the trustee. If it is possible in someway to conclude that the trustee will gain a strong advantage from the situation, we defeasible conclude that he will have a strong commitment to fulfil the expectations.

In general, the trust scheme states that a trustee should not be trusted if it is not clear the benefit for it to perform the task under discussion, and, if the trustee has a strong motivation to achieve a specific task, it could be trusted.

Regarding the computation to be performed, when possible a cost/benefit analysis can be performed. Reviewing this type of computation is not the scope of this work. Cost/benefits analysis is extensively studied. In relation to trust, we could refer to the cost/benefit module of the Secure trust engine [Cah06].

Critical questions to be considered are about the certainness of the computation, and also it there could be extra motivations rather than a potential benefit behind trustee's actions, such as a link of friendship or moral constraints.

Regarding motivations, interesting tests involve the consideration of the reputation value of the trustier, if known. Defeasibly, we could argue that if the trustee has a global reputation to defend, this could be a reasonable supporting argument for motivation. The presence of some rules of behaviour in the environment, or forms of contract is other potential supporter arguments for motivation.

Trust Scheme fulfilment

The trust scheme suggests trusting entities that are committed to fulfil the task assigned. The scheme does not mean that the entity will produce the desiderated outcome, but that he will do its best effort to try to achieve it.

As described by Castelfranchi and Falcone [Cas00], fulfilment means that the trustee will not renounce to achieve the goal, will not search for alternatives, and will pursue the goal for the trustier entity. Fulfilment is a clear sign of commitment and honesty that all are trust ingredients. *Computation*

We propose a computation analogous to the trust scheme past outcomes, in which instead of keeping trace of the number of positive/negative outcomes, we keep trace of the number of

time an entity fulfilled its declared tasks, independently from the outcomes. This means to check how many times the entity completed the task/interactions and the number of time an entity withdrew or could not complete the interactions.

As for the past-outcomes trust scheme, the critical questions involved require to analyse if the lack of fulfilment of an interactions depends from internal or external factors. The discussion performed above can be applied here as well.

Trust scheme derived from exogenous factor: risk and disposition

Trust Scheme Risk

As discussed in the previous chapter, risk is considered in this work a complementary factor to trust. It plays an important role in the decision-making process remaining a separate concept. Therefore, a decision based solely on risk evaluation is not a trust-based decision. Nevertheless, the analysis of risk is an important element that in a complete reasoning can strengthen or weaken conclusions about trust. Here we introduce a trust scheme risk based or the risk profile of trustee and trustier and not on the risk inherent to the situation, that is clearly something external to both the parties and independent from them.

According to the trust scheme risk, the trustier entity A trust the trustee entity B if the risk profile of B is compatible with A's one.

This means that an entity is not inclined to not trust another entity that has a higher attitude to risk. The scheme has a defeasible nature. If the trustee is known to have a very risky attitude the trustier entity should require more evidence to trust him, while in the opposite situation it has no argument against the trustee for what concerns risk.

In order to be applicable the scheme requires knowing the risk profile of the trustee entity to be compared with the trustier's one.

In general this information should be deducted by looking at the risk of the situation where the entities were involved in the past. Regarding the computation of the risk value, this is out of the scope of this thesis about trust, but as described in the previous chapter, in trust studies there are works (the Secure risk module [Cah00]) about risk analysis, based on a classical cost/benefit analysis coupled with past statistical data. If data are available, we could build a risk distribution R(n) showing how the number of members distributed among risk levels and apply a statistical analysis considering the position of a member in respect to other members, and the distance to the centre mass and average value in comparison with the standard deviation too.

An evaluation of the trust scheme risk in the context of a community of online trades is described in the evaluation section. All the reasoning and computation is the same if we consider, instead of the risk profile of the trustee entity, the risk associated to the action/situation. The trustier may not accept too risky situation, independently from the trustee entity.

Appendix D Plausbility, amount of knowledge and aents ranking

This appendix completes the discussion performed in section 4.3.4. Let's consider the following situation. A past-outcome computation has the following strength for each of the agents. Based on this information, the trustier agent A has to derive the conclusion "the trustee agent did well in the past".

- 1. Agent 1: 80%, of good outcomes based on simple average of the marks of 4 exams
- 2. Agent 2: 80% of good outcomes based on a weighted average of 4 exams (difficulty level considered to weight exams' results)
- 3. Agent 2b: 70% of good outcomes based on a weighted average of 4 exams
- 4. Agent 3: 30%, of good outcomes based on simple average of the marks of 4 exams
- 5. Agent 4: 30% of good outcomes based on a weighted average of 4 exams
- 6. Agent 5: 100% of good outcomes based on a weighted average of 4 exams
- 7. Agent 6: 100% of good outcomes based on a weighted average of 8 exams, data reported by the agent and not certified
- 8. Agent 7: 80% of good outcomes based on a weighted average of 4 exams, data reported by the agent and not certified
- 9. Agent 8: 80% of good outcomes based on a weighted average of 8 exams

How could the 8 agents be ordered by Agent A?

First the plausibility base-value $P_{\theta w}$ of a weighted average is greater than the one of a simple average $P_{\theta s}$. The difference is justified by the fact that a weighted average is computed over more information and carries less uncertainty. We remind how this is the default value, that does not consider the effect of external defeater/supporter. Of course, the agent may come to collect evidence showing that its classification of exam difficulties has no plausibility at all, making the weighted average an even weaker argument than a simple average, and it may decide to discard it, but note how this is the effect of an undercutting defeater.

A cautious agent may also decide to put $P_{0w}=P_{0s}$, if he considers that the absence of information about exams' difficulty implies that there is the same probability that the weighted value be worst or better than the simple one.

Again, a key point of this thesis is that an agent must be ready to reason defeasibly, and retract or strengthen its previous conclusions in the light of new evidence.

Therefore, in absence of defeaters, $T_v(Agent 2) > T_v(Agent 1)$ since Agent 2's value is based on a weighted average.

If Agent 2 had 70% of good outcomes, the ordering would depend on the base-value agent A assigned to the two computations. The simple average maybe considered an argument weaker than the weighted one to such an extent that Agent 2 is still trusted more, even with less evidence. In the case of Agent 3 and Agent 4, it is clearly Agent 3 > Agent 4 for similar arguments: they have the same low level of good outcomes, but since Agent 4's results are based on a more plausible computation, its bad results are now amplified.

Agents 5 and 6 have the same trust value if they did the same exams and agent A knows this information, while Agent 6 > Agent 5 in the other case, since $P_{\theta w} > P_{\theta s}$.

The amount of knowledge is expressed by the number of exams available. For instance, Agent 8 has a value computed over more evidence than Agent 2, while the rest is the same. Therefore, Agent 8 > Agent 2.

Finally, Agent 7 is in the same situation as Agent 8, but the exams results were reported by Agent 7 itself, making its value weaker that Agent 8's one.

All the above discussion acquires more emphasis in a dynamic discussion: uncertainty, amount of data, type of computation are all potential counter-arguments that a party (trustee or trustier) can use to rebut the other agent's arguments.

Appendix E Statistical method to isolate confounding variables

In this appendix we describe the statistical method used to isolate the effect of confounding variables.

As described in chapter III and VI, it is important that each trust scheme is computed by considering only relevant data, removing any external dependencies on other variables.

In the description of our method we use as an example the relationship that occurs between the trust scheme activity and longevity, with examples taken from experiment I, FinanzaOnline.it forum.

Between activity and longevity there is a clear correlation that can lead to undesired results. If we need to assess the degree of activity of an entity, we should take in consideration its age as well, since an older entity, even if less active than a younger one, may still have a higher activity. The risk is to measure (and rank) not the degree of activity but actually the age of the entity, that is the confounding variable on which the activity (may) depend.

The idea of our method is to predict the activity that an entity of age X with an activity Y would have if it were of a different age Z. In this way, we can set all the entities to the same age and rank their comparison.

A naive idea could be to use the frequency of the activity to isolate the effect of time. This solution is based on the idea that the entity activity will be constant in the future, and therefore it makes sense to use the frequency of activity to compare activity of entities of different age.

This method is prone to outliers, especially for younger entities, and it does not consider that the dependencies among activity and time is not linear. What makes the difference, especially in the long period, are three variables: the period of inactivity of an entity, the probability that the entity will disappear from the environment, the general degree of activity of the environment. For instance, in *FinanzaOnline.it* experiment, many users that showed an extremely high level of frequency in the first year, are likely to disappear in the following 1-2 years despite this high level. Moreover, the forum shows an increasing level of activity: in the year 2000 an entity on average posted 17% less messages and 40% less attachments.

Our method wants to consider all these factors.

In order to predict the activity of an entity E of age X after a period T>X, our method considers all the older entities that had a similar activity to E, and it considers what these entities did after X-years: their expected lifetime, the expected period of inactivity, the amount of activity they produced after a time T>X.

For instance, if we want to predict the amount of activity of entity E that is 3 year-old and did an activity of 50 so far, we consider what entities that after 3 years showed a similar activity did. There may be 300 similar entities, on which we compute the following:

- their lifetime distribution, from which we can compute the expected value.
- their degree of activity distribution year after year

By using these two distributions, that can be extracted by a statistical analysis of the population, it is possible to predict the expected degree of activity of entity E after T years, value that encompasses the probability that the entity may die.

An uncertainty is attached to the method. The uncertainty depends on:

- the number of entities available to build the distribution of lifetime and activity

- the age of the entity. Young entities have more uncertainty, while older entities have less uncertainty, since their starting activity value is based on more historical information.

The level of uncertainty can be used to further rank entities with similar predicted values, favouring the entities with more certain data. In this way, older entities that have already shown their degree of activity are ranked higher than young entities with similar predicted activity.

In conclusion, the method works as follows. If we need to predict the activity that entity E of age X and activity A would have after a time T>X we need to build:

- 1. the distribution D_L of the lifetime of entities that had an activity close to A ($A \pm e_1$) when they were of age $X \pm e_2$ (e_n is the error that can be tolerated that influences the degree of uncertainty)
- 2. the distribution D_A of entities' activity depending on the age of entity and degree of activity

All the values of activity are normalized using the corrective factors that considers the gereral trend of the forum during the years.

In order to estimate the predicted activity of E, we compute the expected lifetime of E from distribution D_L , and the expected activity of an entity with that average lifetime by looking at D_A .

Entities expected lifetime

The expected lifetime has to be computed in order to assess the probability of an entity to survive in the environment. The expected lifetime of an entity E of age X is computed by considering the lifetime of older entities that had similar characteristic to E. For instance, in the *FinanzaOnline.it* forum, a 3-year old entity with an average degree of activity has an expected lifetime of about 4.5 years.

Corrective Factors

Our analysis of the global activity of the forum showed that in the recent years of the forum entities are keener to post messages and attach files, as if they were more confident with the forum.

On the contrary, in the early years of the forum we register a lower activity per entity, especially for files attachments.

This can lead to a biased situation where younger entities result more active in comparison to older ones. Therefore, we introduced a corrective factor, function of the year where activity was performed, used to normalize the degree of activity. The corrective normalized factors for the action of writing messages and attaching file are shown in the following table:

| Year | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|--|------|------|------|------|------|------|------|------|------|
| Normalized Corrective Factor (messages) | 1 | 1.03 | 1.07 | 1.1 | 1.11 | 1.14 | 1.16 | 1.16 | 1.17 |
| Normalized Corrective Factor (attachments) | 1 | 1.05 | 1.1 | 1.2 | 1.25 | 1.3 | 1.35 | 1.4 | 1.4 |

Appendix F List of Trading Terms used in Competence Analysis

The following is a complete list of trading terms – English and Italian – used to perfom the analysis of competence discussed in Experiment I, chapter VIII.

A REVOCA O CANCELLAZIONE A RISCHIO A VISTA AAA AAA AAA - RATED BOND ABANDON THE CALL / PUT PREMIUM ABBANDONARE IL PREMIO ABOVE PAR ACCANTONAMENTO ACCELERATED DEPRECIATION ACCELERAZIONE DEGLI UTILI ACCEPTANCE ACCERTAMENTO DI IMPOSTA ACCESSO REMOTO ACCETTAZIONE ACCETTAZIONE BANCARIA ACCODAMENTO ACCONTO SUL DIVIDENDO
ACCORDI BANCARI INTERNAZIONALI ACCORDO DI COLLOCAMENTO ACCORDO DI COMPENSAZIONE ACCORDO DI EMISSIONE ACCORDO DI ESTENSIONE ACCORDO DI RINEGOZIAZIONE ACCORDO DI RISTRUTTURAZIONE ACCORDO DI SOTTOSCRIZIONE ACCORDO DI SOTTOSCRIZIONE ACCORDO DI VENDITA ACCORDO FRA SOTTOSCRITTORI ACCORDO GENERALE DI FINANZIAMENTO ACCORDO TRA SOTTOSCRITTORI ACCOUNT ACCOUNT STATEMENT ACCOUNT UNCOLLECTIBLE ACCOUNTS RECEIVABLES
ACCRUED INTEREST ACCUMULATION ACCUMULATION AREA **ACCUMULATION AREA** ACCUMULAZIONE ACCUMULAZIONE ACCUMULAZIONE ACID TEST ACQUIRENTE ACQUISIZIONE CON INDEBITAMENTO ACQUISIZIONE CON INDEBITAMENTO ACQUIST ACQUISTARE ACQUISTO A TERMINE ACQUISTO DI CHIUSURA ACTUAL YIELD ACTUALS AD INTENSITA' DI CAPITALE ADEGUATEZZA PATRIMONIALE

ADR

ADVANCE

ADVERTISING

AFTER HOURS

AGGIOTAGGIO

AGGIUDICAZIONE

AGREEMENT AMONG UNDERWRITERS

AFTRHOURS

AGIOTAGE

AGREEMENT AMONG UNDERWRITERS AIBD AL MEGLIO AL MEGLIO AL MEGLIO AL PORTATORE ALIQUOTA DI IMPOSTA ALL ALL STAR ALLA PARI ALLA PARI ALLENTAMENTO MONETARIO ALLO SCOPERTO ALLOCAZIONE DEGLI INVESTIMENTI ALLOCAZIONE DEGLI INVESTIMENTI ALLOTMENT ALLOTMENT ALPHA AMERICAN DEPOSITARY RECEIPT AMERICAN OPTION AMERICAN STOCK EXCHANGE AMEX
AMMINISTRATORE DELEGATO AMMINISTRATORE DELEGATO, PRESIDENTE AMMORTAMENTO AMMORTAMENTO AMMORTAMENTO A QUOTE COSTANTI AMMORTAMENTO ACCELERATO AMMORTAMENTO PREAMMORTAMENTO **AMORTIZATION** AMORTIZATION ANALISI A QUARTILI ANALISI COSTI BENEFICI ANALISI DEGLI SCENARI ANALISI DELLE REAZIONI ANALISI DI BILANCIO ANALISI DI MERCATO ANALISI DI PORTAFOGLIO ANALISI DI SENSITIVITA ANALISI FONDAMENTALE ANALISI QUALITATIVA ANALISI QUANTITATIVA **ANALISI TECNICA** ANALISI TECNICA ANALISTA ANALYST ANDARE CORTO ANDARE CORTO E LUNGO NEI DERIVATI ANDARE LUNGO ANNO COMMERCIALE ANNO FINANZIARIO ANNOUNCEMENT EFFECT ANNUAL GENERAL MEETING ANNUAL REPORT ANNUALITA' ANNUITY ANNUNCIO SU EMISSIONE DI PRESTITI ANTICIPATION ANTICIPAZIONE ANTICIPAZIONE STRAORDINARIA ANTICIPI E DILAZIONI ANTICIPI E RITARDI ANTICIPO ANTITRUST LAWS APERTO

APERTURA APPRECIATION APPREZZAMENTO APPROCCIO DAL BASSO IN ALTO APPROCCIO DALL'ALTO IN BASSO ARBITRAGE ARBITRAGGIO ARBITRAGGIO COPERTO SU TASSI DI INTERESSE AREA DI ACCOMULAZIONE AREA DI ACCUMULAZIONE AREA DI DISTRIBUZIONE AREA DI DISTRIBUZIONE ARROTONDARE ASK ASPETTATIVE ASSEGNAZIONE ASSEGNO ASSEMBLEA ORDINARIA ASSEMBLEA STRAORDINARIA ASSET ASSET ASSET ALLOCATION ASSET ALLOCATION
ASSET BACKED SECURITIES ASSET BASED LENDING ASSET CLASS ASSET LIABILITY MANAGEMENT ASSET MANAGEMENT ASSET QUALITY ASSET SWAP ASSICURAZIONE ASSOCIAZIONE INTERNAZIONALE PER LO SVILUPPO ASSOCIAZIONE NAZIONALE DEGLI OPERATORI DI BORSA ASSOGESTIONI ASSUNZIONE A FERMO IN BLOCCO ASSUNZIONI A FERMO ASTA ASTA NON COMPETITIVA ASTA OLANDESE AT BEST AT BEST AT PAR AT RISK AT SIGHT AT THE CLOSE AT THE MONEY AT THE MONEY AT THE OPENING ATTIVITA' ATTIVITA' ATTIVITA' CORRENTI ATTIVITA' FISSE ATTIVITA' LIQUIDE ATTIVITA' NETTE ATTIVITA' REALI ATTIVITA' RISCHIOSE ATTRIBUZIONE PRO QUOTA AUCTION AUDIT AUDITING AUMENTO DI CAPITALE AUTOASSICURAZIONE AUTOFINANZIAMENTO AVANZO

BUONI DEL TESORO POLIENNALI BUONO DI SOTTOSCRIZIONE AVERAGE LIFE BENI CAPITALI AVINDEX ® BENI DI CONSUMO AVVERSIONE AL RISCHIO BENI DI CONSUMO DUREVOLI BUONO ORDINARIO DEL TESORO AVVIAMENTO BENI DI CONSUMO NON DUREVOLI **BUTTERFLY SPREAD** AVVISO DI ESERCIZIO BENLEISICI BUY BETA RATIO BUY AZIENDA AZIONE BID BUY / SELL SIGNAL BUY AT BES AZIONE CON DIRITTO DI VOTO BID AZIONE DI COMPENDIO BID / ASK BUY BACK BUY STOP ORDER BIG BOARD BILANCIA COMMERCIALE AZIONE ORDINARIA AZIONE ORDINARIA BUYER AZIONE PRIVILEGIATA DI PRIMA BILANCIA DEI PAGAMENTI BUYING FORWARD BUYING ON MARGIN BUYOUT CATEGORIA AZIONI CALDE AZIONI CICLICHE BILANCIO BILANCIO ANNUALE BILANCIO CONSOLIDATO BUYOUT - TAKE OVER AZIONI DI RISPARMIO **BVPS** BLACK AND SCHOLES BLACK AND SCHOLES MODEL AZIONI NON PRIVILEGIATE AZIONI PRIVILEGIATE CAB CABLE AZIONISTA BLOCCO CAC 40 AZIONISTA DI MAGGIORANZA BLOCK CALENDAR SPREAD BLUE CHIP BOLLA SPECULATIVA AZIONISTA DI RIFERIMENTO AZZARDO MORALE CALL CALL PREMIUM BACK END LOAD BOLLINGER'S BAND CALLABLE CALLABLE BOND CAMBI FLESSIBILI CAMBI SEMIFLESSIBILI BACK TO BACK LOAN BOND BAD DEBT BAD DEBTS BOND BOND CAMBIO BALANCE BOND BOND FUND BOND FUTURE Bond future BOND RATING BALANCE OF PAYMENTS BALANCE SHEET BALANCE SHEET ANALYSIS CAMBIO CAMBIO CAMBIO A TERMINE CAMBIO A TERMINE CAMBIO FISSO CAMBIO FISSO CAMBIO INCROCIATO BALANCED MUTUAL FUND BANCA BOND SWAP BOND WITH WARRANT BOND YIELD BANCA CENTRALE BANCA CENTRALE EUROPEA BANCA DEPOSITARIA BONIFICO CAMBIO INCROCIATO BONIFICO BANCARIO BONUS SHARE BANCA DI INVESTIMENTO BANCA ELETTRONICA CANALE DI TENDENZA CANCELLAZIONE BANCA EUROPEA PER GLI INVESTIMENTI CANCELLAZIONE воок BOOK BOOK RUNNER BOOK VALUE BOOK VALUE BANCA MONDIALE CANDELA BANCA UNIVERSALE BANCABILITA' CANDELE CAP BANCABILITY BOOK VALUE PER SHARE CAPACITA' DI RISCHIO BANCAROTTA
BANCK REAL TIME SECURITIES BORSA DI NEW YORK BORSA FUTURE CAPITAL CAPITAL ADEQUACY DEPARTMENT BORSA MERCI CAPITAL ASSET CAPITAL ASSET CAPITAL GAIN CAPITAL GAIN BORSA SPA BORSA VALORI BANCOMAT BANCONOTA BANDE DI BOLLINGER BORSINO BANDE DI BOLLINGER CAPITAL GAIN TAX CAPITAL GAIN TAX CAPITAL GOODS CAPITAL INCREASE BANDIERA воттом BOTTOM BANDIFRA ВОТТОМ BANK BANK ACCEPTANCE BOTTOM - UP APPROACH CAPITAL INTENSIVE BOUGHT DEAL BOUGHT DEAL CAPITAL LEASE CAPITAL LOSS BANK DRAFT BANK NOTE BANKING SUPERVISION BRAND CAPITAL MARKET CAPITAL MARKET BAR CHART BRAND AWARENESS BRANDING BREAK EVEN BAR CHART CAPITAL REQUIREMENTS BASE BASE DEI PREMI BREAK EVEN CAPITAL TAX BREAK EVEN POINT BREAKOUT BREVE TERMINE BASE MARKET VALUE CAPITALE CAPITALE CIRCOLANTE BASE MONETARIA BASE RATE CAPITALE DI RISCHIO BASIK RISK BREVE TERMINE CAPITALE INVESTITO BASISI POINT BRIDGE LOAN BROKER CAPITALE NETTO CAPITALE NON ANCORA VERSATO
CAPITALE NON SOTTOSCRITTO BROKER BCE CAPITALIZATION CAPITALIZE CAPITALIZZARE BEAR BROKERED DEPOSIT BEAR (ORSO) BEAR MARKET BTF BTP BEAR MARKET BTP FUTURE CAPITALIZZAZIONE CAPOGRUPPO CARTELLO BEAR SPREAD BUDGET BUFFER STOCKS BULL (TORO) BULL AND BEAR BOND BEARER BEFORE TAX PROFIT CARTOLARIZZAZIONE CASH BEI BELOW PAR **BULL MARKET** CASH COLLATERAL BULL MARKET BULL SPREAD BULLET LOAN CASH FLOW CASH FLOW BENCHMARK BENCHMARK BENCHMARK CASH FLOW BENEFICI DI INTERSCAMBIO CASH FLOW YIELD **BUONI DEL TESORO IN ECU** BENEFICIARIO CASH MARKET

CASH MARKET CASH ON DELIVERY CASH RATIO CASH SURRENDER VALUE CASTING VOTE CATEGORIA ASSOGESTIONI CCT CDA CEDOLA CEDOLE (altro significato) CENTRAL BANK CFO CERTIFICATE CERTIFICATE OF DEPOSITE CERTIFICATI DEL TESORO ZERO COUPON CERTIFICATO CERTIFICATO CERTIFICATO AZIONARIO
CERTIFICATO DI DEPOSITO CERTIFICATO OBBLIGAZIONARIO CHANGE CHANGE CHANNEL LINE CHARGE OFF CHECK - CHEQUE CHECKABLE DEPOSITS CHIEDERE CHIEF EXECUTIVE OFFICER CHIEF EXECUTIVE OFFICER CEO CLEAN FLOAT CLEAR CLEARING CLEARING AGREEMENT CLEARING HOUSE CLOSE CLOSED END FUND CLOSING DATE CLOSING PURCHASE CODICE ISIN COEFFICIENTE BETA COLLATERAL COLLEGIO SINDACALE
COLLOCAMENTO
COLLOCAMENTO DIRETTO COLLOCAMENTO PRIVATO COMMERCIAL YEAR
COMMISSION DE SURVEILLANCE COMMISSIONE COMMISSIONE COMMISSIONE DI ENTRATA COMMISSIONE DI GESTIONE COMMISSIONE DI GESTIONE COMMISSIONE DI INCENTIVO O DI PERFORMANCE
COMMISSIONE DI MASSIMO SCOPERTO COMMISSIONE DI PERFORMANCE COMMISSIONE DI RISCATTO COMMISSIONE DI SOTTOSCRIZIONE COMMISSIONE DI SOTTOSCRIZIONE COMMISSIONE DI SOTTOSCRIZIONE COMMISSIONE DI SWITCH COMMISSIONE DI USCITA COMMODITY COMMODITY COMMODITY COMMODITY COMMODITY EXCHANGE COMMODITY FUNDS COMMODITY FUTURE COMMODITY RATES COMMON STOCK COMPARTO COMPENSAZIONE COMPENSAZIONE COMPETITIVE BID COMPETITIVE BID OPTION COMPLIANCE MANAGER

COMPRARE

COMPROMESSO (PRELIMINARE DI CONCENTRAZIONE DI IMPRESA O AZIENDA CONCORDATO FALLIMENTARE CONCORDATO PREVENTIVO CONSEGNA A TERMINE CONSIGLIO DI AMMINISTRAZIONE CONSOB CONSOB CONSOLIDATED FINANCIAL STATEMENT CONSORTIUM CONSORZIO CONSORZIO DI COLLOCAMENTO CONSTANT DOLLAR PLAN CONSUMER CREDIT
CONSUMER DURABLES CONSUMER GOODS CONSUMER PRICE INDEX - CPI CONTANTE CONTANTE CONTINUATION PATTERN CONTO A SALDO NULLO CONTO CORRENTE ORDINARIO CONTO ECONOMICO CONTO FONDO CONTO SOTTOMARGINATO CONTRACT NOTE CONTRATTAZIONE ALLE GRIDA CONTRATTAZIONE CONTINUA CONTRATTAZIONE DI TITOLI NON QUOTATI CONTRATTI IN ESSERE CONTRATTO A CONTANTE CONTRATTO A PREMIO CONTRATTO A TERMINE CONTRATTO APERTO CONTRATTO DI MARGINE CONTRATTO FUTURE CONTRATTO FUTURE SOTTOSTANTE CONTRATTO FUTURE SOTTOSTANTE CONTRATTO OPTION CONTROLLATE CONTROLLO TECNICO CONVENTIONAL LOAN CONVERSION CONVERSION PREMIUM CONVERSION SHARE CONVERSIONE CONVERSIONE CONVERTIBILI CONVERTIBLE BOND CONVERTIBLES CONVERTION RATIO CONVESSIVITA' CONVEXITY COO COOPTAZIONE COORDINATE BANCARIE COPERTURA CORE BUSINESS CORE PORTFOLIO CORPORATE BOND CORPORATE INDENTURE CORPORATION CORRECTION CORRELAZIONE CORREZIONE CORSO SECCO CORSO TEL COST BENEFIT ANALYSIS COST OF CARRY COSTO DI MANTENIMENTO COUNTRY RISK COUPON COUPON STRIPPING COVERED INTEREST ARBITRAGE COVERED OPTION COVERED OPTION **COVERED WARRANT**

COVERED WARRANT COVERED WARRANT CREDIT CRUNCH CREDIT WORTINESS CREDITI CORRENTI CREDITI IN SOFFERENZA CREDITO A RISCHIO CREDITO A RISCHIO CREDITO AL CONSUMO CREDITO APERTO CREDITO APERTO CREDITO COMMERCIALE CREDITO DELLA FEDERAL RESERVE CREDITO INCAGLIATO CREDITO IRREVOCABILE CREDITO REVOLVING CREDITO STAGIONALE CREDITORE CHIROGRAFARO CRESCITA DEI PREZZI CROSS RATE CSSF CTZ CURATORE FALLIMENTARE CURRENT ASSETS CURRENT LIABILITIES CURRENT YIELD CURVA DEI RENDIMENTI CYCLICAL STOCKS DAILY TRADING LIMIT DATA DI CHIUSURA DATA DI ESECUZIONE DATA DI REGOLAMENTO DATA DI SCADENZA DATA EX DIVIDENDE DATE OF MATURITY DAX DAX XETRA DAY LOAN DAY ORDER DEAL STOCK DEALER **DEBITO** DEBITO
DEBITO A BREVE TERMINE
DEBITO CORRENTE DEBITO ESTERNO DEBITO GARANTITO DEBITO NON GARANTITO DEBITO NON GARANTITO DEBITO PRIVILEGIATO DEBITO PRO CAPITE DEBITO PUBBLICO DEBITO SUBORDINATO DERT DEBT INDICATORS DEBT SECURITY DECLASSAMENTO
DEDITO AD AUTORIMBORSO DEFAULT DEFAULT DEFERRED INTEREST DEFICIT COMMERCIALE DEFLATION DEFLATOR DEFLATORE DEFLAZIONE DELISTING DELTA
DEMAND DEPOSIT DENARO DENARO - OFFERTA DEPOSIT DEPOSITARIO DEPOSITARIO DI STATO DEPOSITARY DEPOSITI A VISTA DEPOSITI A VISTA NETTI DEPOSITI IN CONTO CORRENTE DEPOSITO **DEPOSITO DEPOSITO A VISTA**

DEPOSITO BANCARIO DEPOSITO BROKERATO DEPOSITO DERIVATO DEPOSITO ESTERO DEPOSITO FIDUCIARIO DEPOSITO FORWARD FORWARD DEPOSITO INDICIZZATO DEPOSITO INTERBANCARIO **DEPOSITO OVERNIGHT** DEPOSITO TITOLI
DEPOSITO TOM NEXT DEPOSITO VINGOLATO DEPRECIATION DEPREZZAMENTO DEL CAMBIO DEREGOLAMENTAZIONE DEREGULATION DERIVATI DERIVATI LINEARI DERIVATIVE DERIVATIVE DEPOSIT DERIVATIVE INSTRUMENT DERIVATO DERIVATO SU INDICE DI BORSA DERIVATO SU INDICE DI BORSA DESTAGIONALIZZAZIONE DETRAZIONE DEGLI INTERESSI DEVIAZIONE STANDARD DIFFERENTIAL DIFFERENZIALE DIFFERENZIALE DIFFERENZIALE DI EMISSIONE DIFFERENZIALE DI RENDIMENTO DIFFERENZIALE DI SPEZZATURA DIMENSIONE DIRECT INVESTMENTS DIRECT PLACEMENT DIRITTO DI OPZIONE DIRITTO SPECIALE DI PRELIEVO DIRTY FLOAT DISAVANZO COMMERCIALE DISCOUNT DISCOUNT RATE DISCOUNT YIELD DISDETTA DISTRIBUTING SYNDACATE DISTRIBUTION AREA DISTRIBUTION AREA DISTRIBUTORE DISTRIBUZIONE DIVERGENCE DIVERGENZA DIVERSIFICAZIONE DIVIDEND DIVIDEND DISCOUNT MODEL DIVIDEND PAYOUT RATIO DIVIDEND YIELD DIVIDENDI DIVIDENDO DIVIDENDO NON CORRISPOSTO DJIA - DOW JONES INDUSTRIAL AVERAGE DOMESTIC CURRENCY SWAP
DOMICILIAZIONE DELLE BOLLETTE DOMICILIO DEL FONDO DOPPIO MASSIMO DOUBLE TOP DOUDTFUL LOAN DOVERE DI LEALTA' DOW JONES DOW JONES DOW TEORY DOW TEORY DOWN TURN DOWN TURN DOWNGRADING DOWNSIDE RISK DRAFT DROPLOCK BOND **DUAL LISTING** DUE DILIGENCE

DUMPING

DURATA DURATION DURATION MODIFICATA **DUTCH AUCTION** EAD EARNED INCOME EARNING / PRICE RATIO EARNING ASSETS **EARNINGS BEFORE TAXES** EARNINGS MOMENTUM EARNINGS PER SHARE EASDAQ EASY MONEY **EBIT** EBITDA ECCESSO DI RIALZO ECCESSO DI RIALZO (limit high) **ECU** EFFECTIVE DATE EFFECTIVE SALE **EFFETTO ANNUNCIO** EFFETTO GENNAIO EFFICIENCY - X-EFFICIENZA **ELECTRONIC BANKING** ELIOTT EMERGING MARKETS EMISSIONE EMISSIONE EMISSIONE DI AZIONI GRATUITE EMISSIONE SUBORDINATA EMITTENTE EMITTENTE ENTRATA DI CAPITALE ENTRATA NETTA EPS EQUITY EQUITY EQUITY INCOME
EQUITY MARKET
EQUITY MARKET
EQUITY METHOD
EQUITY MUTUAL FUND EQUITY REIT FUNDS **ESEGUITO** ESEGUITO ESENZIONE FISCALE ESERCIZIO ESPOSIZIONE ESTINZIONE O RIMBORSO DEL MUTUO ESTRATTO CONTO ETF ETHIC FUND EUROBBLIGAZIONI EUROCLEAR EUROCURRENCY EURODOLLAR EURODOLLRO EUROMARKET EUROMERCATO EURONOTA EUROPEAN OPTION **EUROVALUTE** EV EV / EBIT EV/SALES EX CEDOLA EX COUPON EX DIRITTI EX DIVIDEND DATE EX DIVIDENDO EX RIGHTS EX WARRANT EXCHANGE DEPRECIATION EXCHANGE RATE EXCHANGE TRADED FUND EXCISE TAX EXERCISE

EXPIRATION EXPOSURE EXPOSURE AT DEFAULT EXTENDED CREDIT **EXTENSION AGREEMENT** EXTERNAL DEBT EXTERNAL FUNDS EXTRAORDINARY GENERAL MEETING FABBRISOGNO DI CAPITALE FACE VALUE FAIR MARKET VALUE FAIR PRICE FAIR VALUE FARE MEDIA FARE MERCATO FASE DI TRADING FASE DI TRADING FASE DI TRENDING FATTURATO FED FEDERAL FUNDS FEDERAL RESERVE - RISERVA FEDERALE USA FIBONACCI FIDEIUSSIONE FIDUCIARIO FIDUCIARY DEPOSIT FIFO FIGURA FIGURE FILL FILL FINANCIAL FUTURE FINANCIAL GUARANTEE FINANCIAL INSTITUTION FINANCIAL LEASE FINANCIAL MARKET FINANCIAL MARKET FINANCIAL POSITION FINANCIAL STATEMENT FINANCIAL YEAR FIRST IN FIRST OUT FIRST PREFERRED STOCK FISSATO BOLLATO FISSAZIONE FIXATION FIXED ASSET FIXED EXCHANGE RATE FIXED EXCHANGE RATE FIXED INCOME INVESTMENTS FIXED LIABILITIES FIXED RATE LOAN FIXING FLAG FLAG FLEXIBLE MUTUAL FUND FLOAT FLOATER FLOATING DEBT FLOATING EXCHANGE RATE FLOATING INTEREST RATE FLOATING RATE BOND FLOATING RATE CERTIFICATE OF DEPOSIT FLOATING RATE NOTE FLOOR FLOOR FLOTTANTE FLUCTUATION FLUCTUATOR **FLUCTUATOR** FLUSSO DI CASSA FLUSSO DI CASSA FLUTTAZIONE FMI FONDAMENTALI FONDI FONDI A COPERTURA FONDI AD ACCUMULAZIONE DEI **PROVENTI**

EXERCISE NOTICE

EXPECTED RETURN

EXPECTATIONS

EXPENSE RATIO

FONDI AGGIUNTI FONDI COMUNI A SCADENZA FONDI COMUNI A SCADENZA FONDI COMUNI DI INVESTIMENTO FONDI DI MATERIE PRIME FONDI ESTERNI FONDI FEDERALI FONDI IMMOBILIARI FONDI NON ACCREDITATI FONDI NON ACCREDITATI FONDI PENSIONE FONDI RETTIFICATIVI FONDI SETTORIALI - SECTOR FUND FONDI STRUTTURALI FONDO FONDO A DISTRIBUZIONE DEI PROVENTI FONDO AD ACCUMULAZIONE DI PROVENTI FONDO CHIUSO FONDO COMUNE APERTO FONDO COMUNE AZIONARIO FONDO COMUNE BILANCIATO FONDO COMUNE FLESSIBILE FONDO COMUNE OBBLIGAZIONARIO FONDO DI AMMORTAMENTO FONDO DI FONDI FONDO DI INVESTIMENTO IMMOBILIARE FONDO DI SETTORE FONDO ETICO FONDO INDICIZZATO FONDO MONETARIO FONDO MONETARIO INTERNAZIONALE -IMF - INTERNETIONAL MONETARY FUND FONDO SENZA COMMISSIONI FORCHETTA DI PREZZO FOREIGN DEPOSIT FOREIGN DIRECT INVESTMENT FOREIGN EXCHANGE FOREIGN MARKET FORWARD FORWARD CONTRACT FORWARD DELIVERY FORWARD EXCHANGE RATE FORWARD EXCHANGE RATE FORWARD MARKET FORWARD MARKET FORWARD PRICE FORWARD PRICE FORWARD SPREAD FORWARD STRUCTURE FORZA RELATIVA FORZA RELATIVA FRAZIONAMENTO FRAZIONAMENTO FRONT END LOAD FTSE FULLY AMORTIZING LOAN FULLY DILUTED EARNINGS **FUND** FUND OF FUNDS FUNDAMENTAL ANALYSIS FUNDING RISK FUNDS FUORI BORSA FUSIONE FUTURE MARKET G.P.F G.P.M. GAIN GAIN FROM TRADE GAP GAP GAPPING RISK GARANZIA GARANZIA GARANZIA DI BUONA ESECUZIONE GARANZIA FINANZIARIA GARANZIA IMMEDIATA GARANZIA IN ORO GARANZIA PERFEZIONATA GARANZIA PRIMARIA

GARANZIA SOTTOSTANTE GARANZIA SOTTOSTANTE GARANZIE AGGIUNTIVE GENERAL AGREEMENT TO BORROW (GAR) GESTIONE ATTIVA GESTIONE DELLE ATTIVITA' / PASSIVITA' GESTIONE DELLE PASSIVITA' GESTIONE PASSIVA GESTORE GIARDINETTO GILT GLOBAL BOND GLOBAL BOND GOING CONCERN VALUE GOING LONG GOING SHORT GOLD BACKING GOLD INDEXED INVESTMENT GOLD STANDARD GOLDEN SHARE GOOD THIS MONTH GOOD TILL CANCELLED GOODWILL GOVERNMENT BOND GOVERNMENT DEPOSITORY GOVERNMENT NATIONAL MORTGAGE ASSOCIATION GPF - GESTIONI PATRIMONIALI FONDI GPM - GESTIONE PATRIMONIO MOBILIARE GRAFICO GRAFICO A BARRE GRAFICO A BARRE GRAFICO LINEARE GRAFICO LINEARE GRAY MARKET GREEN SHOE GREENSHOE GREY MARKET GRIDA GROSS PROFIT GROSS YIELD **GROWTH STOCK** GRUPPO GRUPPO DI DIREZIONE GRUPPO DI MUTUI IPOTECARI GRUPPO DI SOTTOSCRIZIONE GRUPPO DI SOTTOSCRIZIONE GRUPPO DI SOTTOSCRIZIONE A FERMO GRUPPO DI VENDITORI GUADAGNO IN CONTO CAPITALE GUARANTEE HANG SENG HEAD & SHOULDERS HEAD & SHOULDERS HEDGE HEDGING HISTORICAL VOLATILITY HKIBOR HOLD HOLDER HOLDING I.S.C. IMMOBILIZZAZIONI IMMOBILIZZAZIONI MATERIALI IMMOBILIZZAZIONI PRODUTTIVE IMPLICIT INTEREST RATE IMPLIED VOLATILITY
IMPOSTA DI REGISTRO IMPOSTA IN CONTO CAPITALE IMPOSTA PROGRESSIVA IMPOSTA SU TITOLI ESTERI IMPOSTA SUI BENI VOLUTTARI IMPOSTA SUL MONTE SALARI IMPRESA DI PUBBLICA UTILITA' IN - FEE IN APERTURA IN CHIUSURA IN THE MONEY

INDEBITAMENTO NETTO INDEX BOND INDEX DOW JONES INDEX FUND INDEX S&P INDEX STOCK INDEXED FUND INDICATORE
INDICATORE ANTICIPATORE INDICATORI DI ASPETTATIVA INDICATORI DI ASPETTATIVE INDICATORI DI INDEBITAMENTO INDICATORI DI REDDITIVITA' INDICE INDICE AZIONARIO INDICE BANCA DI ITALIA INDICE DEI PREZZI AL CONSUMO INDICE DEI PREZZI ALLA PRODUZIONE INDICE DEI PREZZI ALL'INGROSSO INDICE DI SHARPE INDICE NIKKEI INDICE SOTTOSTANTE INDICIZZATO INDICIZZAZIONE INFORMATION RATIO INITIAL MARGIN INSIDER TRADING INSOLVENZA INSTITUTIONAL INVESTOR INSTITUZIONALI INSURANCE INTERBANK DEPOSIT INTERBANK RATE INTERBANK RATE INTERESSE INTERESSI DI MINORANZA INTERESSI DIFFERITI INTEREST INTEREST EQUALIZATION TAX INTEREST RATE INTEREST RATE CAP INTEREST RATE FUTURE INTEREST RATE SWAP TARGET PRICE STOP LOSS
INTERIM DIVIDEND
INTERIM RESULTS BUSINESS PLAN FUTURES INTERMEDIARIO INTERMEDIARIO INTERNAL DEALING INTERNATIONAL BANKING FACILITIES INTERNATIONAL DEVELOPMENT ASSOCIATION (IDA) INTERVALLO INTERVALLO DI CONTRATTAZIONE INTERVALLO DI OSCILLAZIONE INTERVALLO DI OSCILLAZIONE INTRINSIC VALUE INVERSIONE INVERSIONE INVESTED CAPITAL INVESTIMENTI A CAPITALE COSTANTE INVESTIMENTI A REDDITO FISSO INVESTIMENTI DIRETTI INVESTIMENTI FACILI DA LIQUIDARE INVESTIMENTI MATERIALI PER AZIONE INVESTIMENTO
INVESTIMENTO DI CAPITALE INVESTIMENTO IN SOFFERENZA INVESTIMENTO INDICIZZATO ALL'ORO INVESTIMENTO STRANIERO DIRETTO INVESTITORE ISTITUZIONALE INVESTMENT INVESTMENT BANK INVESTMENT CASE INVESTMENT LETTER INVESTMENT RISK INVESTOR RELATIONS INVIM

MARKET OUTPERFORM MARKET PERFORM LIMITE SUPERIORE / INFERIORE **IPERVENDUTO IPERVENDUTO** LINE CHART MARKET PRICE IPO LINE CHART LINEA DI TENDENZA MARKET RISK MARKET RISK IPOTECA IPOTECA DI PRIMO GRADO LINEAR DERIVATIVES MARKET SHARE IRR LIQUID ASSETS MARKET TIMING MARKET UNDERPERFORM MARKET VALUE IRS LIQUIDABILITA'
LIQUIDATINE VALUE ISC - INDICATORE SINTETICO DEL COSTO LIQUIDATION MARKETABILITY ISIN CODE ISLAND REVERSAL LIQUIDAZIONE MASSIMI LIQUIDITA' LIQUIDITY MASSIMO ISSUE MATERIA PRIMA LIQUIDITY RISK LISTA DI CONTROLLO ISSUE MATERIE PRIME ISSUE PRICE MATIF MATURITA' MATURITY LISTING REQUIREMENTS LISTINO UFFICIALE ISSUER ISSUER ISSUING SYNDACATE LIVELLO MAX MEDIA MEDIA MOBILE MEDIA MOBILE ISTITUZIONE FINANZIARIA LOAD JANUARY EFFECT JENSEN ALPHA LOAD LOAN JOINT VENTURE LOAN GRADING MEDIA PONDERATA MEDIA STORICA MERCATI A PRONTI MERCATI EMERGENTI JUNIOR STOCK LOCAZIONE JUNK BOND JUNK BOND LOCAZIONE A LEVA LOCAZIONE APERTA JUNK BOND BUST UP TAKEOVER LOCAZIONE DI CAPITALE MERCATO LOCAZIONE FINANZIARIA LOCAZIONE FINANZIARIA MERCATO MERCATO A PRONTI JUNK BONDS JUST IN TIME KAFFIRS LOCK UP MERCATO AFTERHOUR MERCATO AZIONARIO MERCATO AZIONARIO MERCATO DEI CAMBI KENNEDY ROUND LOCK UP PERIOD KEY CURRENCY KEY RATE LOMBARD RATE LONG KEY REVERSAL DAY LONG POSITION MERCATO DEI CAMBI KICK LONG POSITION MERCATO DEI CAPITALI KICKBACK KURTOSIS MERCATO DEI CAPITALI MERCATO DEL TERMINE MERCATO DEL TERMINE LONG TERM LOSS LOT MERCATO DELLE SPEZZATURE MERCATO DELLE SPEZZATURE LABOUR UTILIZATION RATE LOTTI MINIMI / LOTTO MINIMO LADDERED PORTFOLIO LARGA CAPITALIZZAZIONE LOTTO LOTTO ROTONDO MERCATO FINANZIARIO LARGE CAPS MERCATO FINANZIARIO LOW LAST IN FIRST OUT LAST SALE MERCATO GRIGIO MERCATO GRIGIO LOW LUNGO TERMINE LAST TRADING DAY M&A MERCATO MONETARIO MERCATO MONETARIO MERCATO NON REGOLAMENTATO MERCATO OBBLIGAZIONARIO LATE TAPE M1 IBO M2 LEADING INDICATORS МЗ MERCATO OBBLIGAZI MERCATO ORSO MERCATO PRIMARIO MERCATO PRIMARIO LEADS AND LAGS LEADS AND LEGS LEASE - PLACE MACROECONOMIA MACROECONOMICS LEASING MAGAZZINO MAGAZZINO MAJORITY SHAREHOLDER MAKE AVERAGE MERCATO RISTRETTO MERCATO RISTRETTO MERCATO SECONDARIO LEGAL RESERVES LENDER LENDER OF LAST RESORT LETTER OF CREDIT MAKE MARKET MERCATO SECONDARIO MERCATO TORO
MERCATO TORO LETTERA MANAGED CURRENCY LETTERA DI CREDITO LETTERA DI INVESTIMENTO MANAGED LIABILITIES MANAGEMENT FEE MERCATO VALUTARIO LETTERA SUL MERCATO MANAGEMENT GROUP MERCHANT BANK LEVA FINANZIARIA MANI FORTI MERCI MARGIN MERCI LEVERAGE LEVERAGE MARGIN AGREEMENT MERGER LEVERAGE BUYOUT MARGINE MERGER AND ACQUISITION LEVERAGE BUYOUT LEVERAGED LEASE MARGINE MARGINE DI VARIANZA MARGINE DI VARIAZIONE MERITO DI CREDITO METODO DEL PATRIMONIO NETTO MEZZI PROPRI LIABILITY LIABILITY MANAGEMENT MARGINE INIZIALE MIB MARGINE OPERATIVO MARGINE OPERATIVO LORDO MIB 30 LIBID MIBTEL LIBOR MARK TO MARKET MICROECONOMIA LIBRETTO DI RISPARMIO LIBRO MARKET MICROECONOMICS LIBRO NON COPERTO MARKET MID CAP LIFE ANNUITY MARKET ANALYSIS MID CAPS LIFE INSURANCE IN FORCE MARKET INDEX DEPOSIT MIDEX MARKET LETTER
MARKET LIQUIDITY RISK
MARKET MAKER
MARKET MAKER LIFFE MIF MINIMI LIFO LIFT MINIMO MINIMO INCREMENTO LIMIT HIGH MARKET MAKING MINIMO INCREMENTO MARKET NEUTRAL MINIMUN LOT LIMIT UP / DOWN

LIMITE DI OSCILLAZIONE

MARKET ORDER

IPERCOMPRATO IPERCOMPRATO

MINORITY INTEREST MINUSVALENZA MIVING AVERAGE MODEL MODELLO MODELLO DI BLACK & SCHOLES MODELLO DI CONSOLIDAMENTO DI CONTINUAZIONE MODELLO DI CONSOLIDAMENTO E DI CONTINUAZIONE MODELLO DI INVERSIONE MODELLO DI INVERSIONE MODELLO DI PREZZO MODELLO DI SCONTO DEI DIVIDENTI MODELO DI PREZZO MODIFIED DURATION MOL MOMENTO MOMENTUM MONETA MONETARY BASE MONETARY POLICY MONEY MONEY MANAGER MONEY MARKET MONEY MARKET MONEY MARKET FUND MONEY MARKET RATES MONEY SUPPLY MOODY'S MORAL HAZARD MORTGAGE POOL MOT MOVING AVERAGE MSCI MTA МТО MTS MULTIPLI MUTUO MUTUO FONDIARIO NAFTA NAIC NAIRU NAKED OPTION NAKED POSITION NAKED POSITION NAME NASD NASDAQ NASDAQ 100 NASDAQ Composite Index NASDAQ Small Cap Market NASTRO DEI PREZZI NATIONAL ASSOCIATION OF INVESTOR CORPORATION NATIONAL ASSOCIATION OF SECURITIES DEALERS (NASD) NAV NEGOTATION NEGOTIABLE NEGOZIABILE NEOLIBERISMO NET ASSET VALUE NET ASSET VALUE NET ASSETS NET BORROWED **NET CHANGE** NET DEBT NET DEMAND DEPOSIT NET FINANCIAL POSITION NET INCOME NET INCOME NET INCOME PER SHARE NET INTEREST INCOME PER SHARE NET PRESENT VALUE NET PROCEEDS NET TANGIBLE ASSET PER SHARE

NET WORTH

NET WORTH

NET YIELD NETTING NETTO NEW ECONOMY NEW ISSUE NEW MONEY NEW YORK STOCK EXCHANGE - NYSE NICCHIA NICHE NIF NIKKEI 225 NIKKEI INDEX NO LOAD NO LOAD FUND NO PERFORMING LOAN NO REFUNDABLE NOME NOMINAL INTEREST RATE NOMINALE NOMINALE DI RIMBORSO NON NON CALLABLE NON COMPETITIVE BID NON DURABLE GOODS NON REGOLAMENTATO (in riferimento al mercato) NON RINNOVABILE NON TRASFERIBILE NONACCRUAL ASSET NORMAL TRADING UNIT O ROUND LOT NORME ANTITRUST NOTA NOTA A TASSO VARIABILE NOTA DI PAGAMENTO NOTE NOTE INDICIZZATE NOTE ISSUANCE FACILITY NOTE NOTICE NOTES NOW NUMTEL NUOVA ECONOMIA NUOVA EMISSIONE NUOVO MERCATO NYBOR NYSE O.B.V. - ON BALANCE VOLUNE O.P.A. OBBLIGAZIONE OBBLIGAZIONE OBBLIGAZIONE **OBBLIGAZIONE** OBBLIGAZIONE A GARANZIA FISCALE ILLIMITATA OBBLIGAZIONE A TASSO VARIABILE
OBBLIGAZIONE CONVERTIBILE OBBLIGAZIONE DROPLOCK OBBLIGAZIONE INDICIZZATA OBBLIGAZIONE SENZA CEDOLE OBBLIGAZIONE STRUTTURATA OBBLIGAZIONI OBBLIGAZIONI A MEDIO TERMINE OBBLIGAZIONI A TASSO DI INTERESSE OBBLIGAZIONI A TASSO FISSO OBBLIGAZIONI BULL AND BEAR OBBLIGAZIONI CONVERTIBILI OBBLIGAZIONI CONVERTIBILI A CEDOLA NULLA OBBLIGAZIONI DELLE SOCIETA' OBBLIGAZIONI DOMESTICHE OBBLIGAZIONI ESTERE OBBLIGAZIONI IN EURO - EUROBOND OBBLIGAZIONI NON INVESTMENT GRADE OBBLIGAZIONI SOCIETARIE OBBLIGAZIONI ZERO COUPON OBV ODDIOT ODD LOT MARKET ODD LOT MARKET OFF

OFF - SPIN OFF OFF MEETING OFFERING CIRCULAR OFFERTA OFFERTA COMPETITIVA OFFERTA COMPETITIVA CON OPZIONI OFFERTA DI ACQUISTO OFFERTA DI ACQUISTO OFFERTA DI GARANZIA OFFERTA DI MONETA OFFERTA GLOBALE OFFERTA GLOBALE
OFFERTA PUBBLICA
OFFERTA VISIBILE
OFFICIAL LIST
OFFICIAL PRICE OFFICIAL RESERVES ONDE ONERI ACCESSORI OPAS OPAS OPEN OPEN AND LEASE OPEN CONTRACT **OPEN CREDIT** OPEN END CREDIT OPEN END FUND OPEN INTEREST OPEN ORDER OPEN OUTCRY OPEN PRICE BOOKBUILDING OPENING OPERATING MARGIN OPERATING PROFIT OPERATING PROFIT OPERATING RESULTS OPERATORE **OPERATORE** OPS OPTION SPREAD OPTION WRITER OPTIONS CLEARING CORPORATION OPV OPVS OPZIONE OPZIONE AMERICANA OPZIONE AMERICANA OPZIONE AT THE MONEY OPZIONE CALL OPZIONE CALL OPZIONE COPERTA OPZIONE DI ACQUISTO OPZIONE DI SWAP OPZIONE DI VENDITA OPZIONE DOPPIA OPZIONE EUROPEA OPZIONE EUROPEA
OPZIONE IN THE MONEY OPZIONE OUT OF THE MONEY OPZIONE PUT
OPZIONE SCOPERTA
OPZIONE SCOPERTA OPZIONE SCOT ENT OPZIONI - OPTION OPZIONI SU AZIONI ORA DELLE STREGHE ORDER ORDINARY SHARE ORDINE ORDINE AL MEGLIO ORDINE APERTO ORDINE DI ACQUISTO ORDINE DI ACQUISTO CON STOP ORDINE DI CHIUSURA IN PERDITA ORIZZONTE PERIODALE ORSO OSCILLATORE OSCILLATORE
OSCILLATORE STOCASTICO OSCILLATORI OTC

OTC MARKET PREZZO UFFICIALE OUT OF THE MONEY OVER THE COUNTER - OTC POINT OF SALE
POLITICA MONETARIA PREZZO UFFICIALE PREZZO UTILE PRICE / BOOK VALUE PRICE / EARNING OVER THE COUNTER MARKET POLIZZE VITA IN ESSERE MERCATO NON REGOLAMENTATO PORTAFOGLIO PORTAFOGLIO **OVERBORROWING** PRICE / EARNING RATIO **OVERBOUGHT** PRICE CHANGE OVERBOUGHT PORTAFOGLIO A GRADINI PRICE EARNING OVERCOLLATERALIZATION PORTAFOGLIO BARBELL PRICE PATTERN PRICE PATTERN PRICE SPREAD OVERDRAFT PORTAFOGLIO RACCOMANDATO OVERI ENDING PORTFOLIO OVERNIGHT DEPOSIT PORTFOLIO ANALYSIS PRICING OVERNIGHT POSITION POS PRIMARY MARKET POSITION PRIMARY MARKET
PRIMARY RESERVES **OVERNIGHT RATE OVERSOLD** POSIZIONE
POSIZIONE CORTA **OVERSOLD** PRIME RATE OVERWRITING POSIZIONE CORTA PRIME RATE POSIZIONE FINANZIARIA POSIZIONE FINANZIARIA NETTA PRINCIPAL AMOUNT PRINCIPAL SHAREHOLDER P/BOOK P / E + A P/U POSIZIONE LUNGA PRIOR LIEN PAC POSIZIONE LUNGA PRIVATE EQUITY PAESI EMERGENTI PAGAMENTO ALLA CONSEGNA PRIVATE PLACEMENT PRODUCER PRICE INDEX - PPI POSIZIONE OVERNIGHT POSIZIONE OVERNIGHT PAGAMENTO CONTRO PAGAMENTO POSIZIONE SCOPERTA PROFIT PAGAMENTO IN CONTANTI POSIZIONE SCOPERTA PROFIT AND LOSS ACCOUNT PROFIT TAKING PROFIT TAKING PAGAMENTO PERIODICO PAGHERO' CAMBIARIO POTENZIALE AL RIALZO PRE PANIC BUYING PREAPERTURA PROFIT WARNING PRE-APERTURA
PREFERENCE SHARE PANIC SELLING PROFITABILITY PROFITABILITY INDICATORS PAR PROFITTO PARADISO FISCALE PREMIO PROFITTO NON REALIZZATO PARAMETRO DI INDICIZZAZIONE PREMIO AL RISCHIO PREMIO DI CONVERSIONE PREMIO DONT PROGRAM TRADING PROGRESSIVE TAX PARCO BUOI PARI PASSIVITA' PREMIO NON AMMORTIZZATO PRONTI CONTRO TERMINE PASSIVITA' CORRENTI PREMIUM PROPENSIONE AL RISCHIO PASSIVITA' FISSE PASSIVITA' GESTITE PRENDERE PROFITTO PREPAYMENT PROSPETTO INFORMATIVO PROSPETTO INFORMATIVO 2 PATRIMONIO PRESA DI BENEFICIO PROVISION PROVISION
PROXY VARIABLE
PUBLIC DEBT
PUBLIC OFFERING
PUBLIC UTILITIES PRESA DI BENEFICIO PRESENT VALUE PATTO DI SINDACATO **PAVIMENTO** PRESTATORE PAYEE PAYMENT VERSUS PAYMENT PRESTATORE DI ULTIMA ISTANZA PAYOUT DEI DIVIDENDI PAYOUT RATIO PRESTITI GARANTITI PRESTITI IN SOSPESO PUNTO BASE PURCHASE ORDER PAYOUT RATIO **PRESTITO** PUT PRESTITO A TASSO FISSO PRESTITO AD AMMORTAMENTO PAYROLL TAX PUT OPTION PBV Q RATIO COMPLETO QUALIFYING SHARE PEG PENALE ESTINZIONE ANTICIPATA PRESTITO BACK TO BACK QUALITA' DEGLI INVESTIMENTI PENNANT PENNELLO PRESTITO BOW PRESTITO BULLET QUALITATIVE ANALYSIS QUANTITÀ PRESTITO CONSORZIATO QUANTITATIVE ANALYSIS PENSION FUNDS PER CAPITA DEBT PRESTITO DI TITOLI QUARTERLY REPORT (RELAZIONE PRESTITO GARANTITO
PRESTITO GARANTITO CONVENZIONALE PERDITA TRIMESTRALE) PERDITA IN CONTO CAPITALE QUEL PERFECTED LIEN PRESTITO GARANTITO DA DEPOSITO QUEUING QUICK RATIO QUIETANZA DI PAGAMENTO PERFORMANCE PRESTITO GIORNALIERO PRESTITO NON GARANTITO PRESTITO NON GARANTITO PERFORMANCE PERFORMANCE BOND QUORUM PERIODO DI BLACK OUT PRESTITO PERSONALE QUOTA DI FONDO PRESTITO PONTE
PRESTITO RINEGOZIATO QUOTA DI MERCATO QUOTA PARTE PERIODO DI CHIUSURA PFRIZIA PRESTITO RINNOVABILE QUOTA QUALIFICANTE PERSONAL LOAN PIANO DI ACCUMULO DEL CAPITALE e PRESTITO SINDACATO QUOTAZIONE QUOTAZIONE (QUOTE) QUOTAZIONE CERTO PER INCERTO PRESTITO STAGIONATO PIANO DI AMMORTAMENTO PRESTITO SU TITOLI PRESTITO TENTENNANTE PIANO DI INVESTIMENTO QUOTE PIANO DI RIMBORSO PROGRAMMATO PRESTITO TENTENNANTE QUOZIENTE DI INDEBITAMENTO PREZZO A PRONTI PREZZO A TERMINE QUOZIENTE DI LIQUIDITÀ QUOZIENTE DI TESORERIA PIATTINO PIATTINO PREZZO CONTANTE QUOZIENTE Q PIAZZA AFFARI PREZZO DENARO R QUADRO PREZZO DI APERTURA PREZZO DI EMISSIONE RAGGRUPPAMENTO PIN RALLY PIVOT PREZZO DI ESERCIZIO RALLY PIVOT PLACEMENT PREZZO DI ESERCIZIO STRIKE PRICE RANGE PREZZO DI MERCATO PREZZO DI REGOLAMENTO PLAIN VANILLA SWAP RANGE RAPM PLEDGE PREZZO DI RIFERIMENTO RAPPORTO ANNUALE PLUSVALENZA PLUSVALENZA PREZZO EQUO RAPPORTO DI CONVERSIONE

RAPPORTO DI SPESA RAPPORTO DI TREYNOR RAPPORTO PREZZO / UTILE RAPPORTO PREZZO UTILE RAPPORTO SEMESTRALE RAPPORTO UTILI / PREZZO RATA RATE RATE SENSITIVE RATEO RATEO DI INTERESSE RATING RATING DELLE OBBLIGAZIONI RATIOS. Valori di sintesi che forniscono indicazioni sullo stato di salute di una impresa, tra cui i più importanti sono REAL ASSETS
REAL ESTATE INVESTMENT TRUST REAL RATE OF RETURN REALE REBOUND RECCOMENDED LIST RECEIVE VERSUS PAYMENT RECOVERY RATIO RECOVERY RAYTE RECURRING PAYMENT REDDITIVITA' REDDITIVITA' DEL CAPITALE AL VALORE DI MERCATO REDDITO FISSO REDDITO NETTO REDDITO OPERATIVO REDEMPTION REDUCE REGIME DEL RISPARMIO AMMINISTRATO REGIME DEL RISPARMIO GESTITO
REGOLA DELL'INCREMENTO REGOLA DELL'INCREMENTO REGOLA DELLO SCOPERTO REGOLAMENTO REGOLAMENTO DEL FONDO REGOLAMENTO DEL FONDO RELATIVE STRENGHT RELATIVE STRENGHT RELATIVE STRENGTH INDEX REMOTE ACCESS RENDIMENTO RENDIMENTO RENDIMENTO RENDIMENTO A SCADENZA RENDIMENTO A SCONTO RENDIMENTO ALLA SCADENZA RENDIMENTO ATTESO RENDIMENTO DEI MEZZI PROPRI RENDIMENTO DEL CAPITALE INVESTITO RENDIMENTO DI CASSA RENDIMENTO DI UNA OBBLIGAZIONE RENDIMENTO EFFETTIVO RENDIMENTO IMMEDIATO RENDIMENTO LORDO RENDIMENTO NETTO RENDIMENTO REALE RENDIMENTO TOTALE RENDIOB RENDISTATO RENDITA VITALIZIA RENEGOTIATED LOAN REQUISITI DI RISERVA REQUISITI PER LA QUOTAZIONE RESCHEDULING AGREEMENT RESIDENZA RESIDUAL CLAIMANT RESIDUAL VALUE RESISTANCE RESISTENZA RESISTENZA RETAINED EARNINGS RETENTION

RAPPORTO DI DISTRIBUZIONE

RETRACEMENT RETURN Return On Capital RETURN ON EQUITY RETURN ON INVESTMENT RETURN ON SALES REVERSAL REVERSAL REVERSAL PATTERN REVERSAL PATTERN REVERSE REVERSE CONVERTIBLE REVERSE FLOATER REVERSE HEAD AND SHOULDERS REVERSE REQUIREMENTS REVERSE SPLIT REVISIONE CONTABILE REVOLVING CREDIT RIACQUISTARE - ACQUISTO DI AZIONI PROPRIE RIACQUISTO E ACQUISTO DI AZIONI PROPRIE RIALZISTA RIALZO RIALZO RIBASSISTA RIBASSO RIBASSO RIBOR RICAPITALIZZARE RICAPITALIZZAZIONE RICAVO RIDURRE RIMBALZO RIMBORSABILE RIMBORSO RIMBORSO ANTICIPATO RINNOVO RINTRACCIAMENTO RIPARTIZIONE RIPORTO - LENDING SECURITIES RISCHI DI ASIMMETRIA DELLE SCADENZE RISCHIO RISCHIO AL RIBASSO RISCHIO DEL CAPITALE RISCHIO DI BASE RISCHIO DI FINANZIAMENTO RISCHIO DI INVESTIMENTO RISCHIO DI LIQUIDITA' RISCHIO DI LIQUIDITA' DEL MERCATO RISCHIO DI MERCATO RISCHIO DI MERCATO RISCHIO PAESE RISCHIO SPECIFICO RISCHIO STANDARD RISERVA riservati. Per info RISERVE OBBLIGATORIE RISERVE PRIMARIE RISERVE SECONDARIE RISERVE TOTALI RISERVE UFFICIALI RISK RISK ADJUSTED PERFORMANCE MEASURE RISK ASSETS RISK AVERSION RISK BASED CAPITAL RISK CAPITAL RISK FREE RATE RISK PREMIUM RISORSA SOGGETTA AD ESAURIMENTO RISPARMIO RISTRUTTURAZIONE RISULTATO OPERATIVO RITARDO NEI PREZZI RITENUTA FISCALE RITENZIONE RITORNO

ROAD SHOW ROCA ROE ROI ROLLOVER ROS ROUND LOT RSI S & P S&P 500 SAFE KEEPING SAL SALDO SALDO CONTABILE SALDO DISPONIBILE SALDO LIQUIDO SALE SALE AND LEASEBACK SALES REVENUE SAUCER SAUCER SAVING ACCOUNT LOAN SAVINGS SAVINGS SHARE SCADENZA SCADENZA DELLA OPZIONE SCALATA SCALATA CON DISGREGAZIONE PER PAGARE I JUNK BOND SCAMBIO SCARICO SCARPA VERDE SCARTO DI EMISSIONE SCENARIO ANALYSIS SCONTO SCONTO NON AMMORTIZZATO SCOPERTO SCORING SCORTE CUSCINETTO SCRIP SEASONAL ADJUSTMENT SEASONAL CREDIT SEASONED LOAN SEC SECONDARY IPO SECONDARY MARKET SECONDARY MARKET SECONDARY RESERVES SECTOR SECURED LOAN SECURITIES LOAN SECURITIES UNDERWRITING SECURITY SECURITY INTEREST SECURITY RATING SECURIZATION SEGNALE DI ACQUISTO O DI VENDITA SELECTED DEALER AGREEMENT SELF INSURANCE SELF SUPPORTING DEBT SELF-FINANCING SELL SELL DOWN SELL OFF SELL PLUS SELL SHORT SELL THE BOOK SELLING AGREEMENT SELLING GROUP MEMBERS SEMESTRALE SENIOR DEBT SENSIBILE AI TASSI SENSITIVITY ANALYSIS SENTIMENT SENTIMENT INDICATOR SENTIMENT INDICATOR SENTIMENTO SENZA COMMISSIONI SENZA RIMBORSO ANTICIPATO SETTLEMENT SETTLEMENT DATE

RITRACCIAMENTO

ROA

TASSI INTERBANCARI TASSI SU MATERIE PRIME SETTLEMENT PRICE SPREAD RIALZISTA SPREAD RIBASSISTA SETTORE SEVERAL BUT NOT JOINTLY SPREAD TEMPORALE TASSO SGR STAGFLAZIONE TASSO A REGIME STANDARD & POOR'S Standard & Poor's 500 STANDARD & POOR'S INDICE SHARE TASSO CHIAVE TASSO DI CAMBIO FLESSIBILE SHARE O STOCK SHARE PREMIUM TASSO DI INTERESSE SHAREHOLDER O STOCKOLDER STANDARD RISK TASSO DI INTERESSE STANDARDIZED APPROCH SHARPE TASSO DI INTERESSE CAP TASSO DI INTERESSE IMPLICITO SHARPE INDEX STANZA DI COMPENSAZIONE SHARPE RATIO STANZA DI COMPENSAZIONE PER LE TASSO DI INTERESSE INDICIZATO SHORT **OPZIONI** TASSO DI INTERESSE NOMINALE STATO PATRIMONIALE TASSO DI RECUPERO DEL CREDITO SHORT TASSO DI RIFERIMENTO SHORT POSITION STOCASTIC SHORT POSITION STOCASTICO INDICATORE TASSO DI SCONTO SHORT SALE RULE SHORT SELL SHORT TERM SHORT TERM STOCHASTIC INDICATOR TASSO D'INFLAZIONE STOCK TASSO FISSO TASSO INTERNO DI RENDIMENTO STOCK BUYBACK STOCK EXCHANGE TASSO MISTO TASSO OVERNIGHT TASSO PRIMARIO SHORT TERM DEBT SHORT TERM INTEREST RATE STOCK INDEX DERIVATIVE STOCK INDEX DERIVATIVE STOCK OPTION TASSO RISK FREE SICAV SICURED DEBT STOCK PICKING TASSO SULLE ANTICIPAZIONI STOCK SELECTION STOCK SPLIT SIGHT DEPOSIT TASSO UFFICIALE DI SCONTO TASSO VARIABILE SIGHT DRAFT Sigla di London Interbank Offered Rate STOP AND GO TAX ASSESSMENT STOP LOSS TAX EXEMPT SINDACATO DI DISTRIBUZIONE SINDACATO DI EMISSIONE STOP LOSS ORDER STOP PROFIT TAX HEAVEN TAX RATE SINDACAZIONE STORE RETAILING TAX SELLING SINKING FUND STOXX 50 TECH CHECK SISTEMA DELLA FEDERAL RESERVE TECHNICAL ANALYSIS TECHNICAL ANALYSIS STRADDLE STRAIGHT LINE DEPRECATION SIZE SOCIETA' CONTROLLATA STRANGLE TELECOMUNICATIONS, MEDIA AND SOCIETA' DI INVESTIMENTO A CAPITALE STRATEGIA BOTTOM UP **TECHNOLOGIES** VARIABILE - OPEN ENDED INVESTMENT STRATEGIA TOP DOWN TENDENZA COMPANY STRIKE PRICE TENDENZA SOCIETA' DI INVESTIMENTO MOBILIARE STRONG BUY TENDER SOCIETA' GESTIONE RISPARMIO SOCIETA' PER AZIONI STRONG CURRENCY STRONG SELL TENDER OFFER TENDER OFFER SOPRA LA PARI STRUCTURAL FUNDS TEORIA DI DOW SOSPENSIONE STRUMENTO DERIVATO TEORIA DI DOW SOTTO LA PARI SOTTOCAPITALIZZAZIONE STRUTTURA A TERMINE SUBORDINATED DEBT TERM TERMINE TESTA E SPALLA ROVESCIATO TESTA E SPALLE TESTA E SPALLE SOTTOSCRITTORE SUPPORT SUPPORTO SUPPORTO SOTTOSCRIZIONE SOTTOSCRIZIONE SOTTOSCRIZIONE DI TITOLI SURPLUS TESTA E SPALLE SOTTOSTANTE SURPLUS TICKER SOTTOSTANTE SOTTOVALUTATO TIE - BOW - TIE LOAN SWAP TIME DEPOSIT SWAP SOVRAPPREZZO AZIONI SWAP DI ATTIVITA' TIME DRAFT SPECIAL DRAWING RIGHT SPECIAL OFFERING SWAP DI OBBLIGAZIONE SWAP DI TASSO DI INTERESSE TIME SPREAD TIME VALUE **SPECIALIST** SWAP VALUTARIO DOMESTICO TITOLARIZZAZIONE SPECIALISTA SWAPTION TITOLI TITOLI AL PORTATORE
TITOLI CARTOLARIZZATI IPOTECAR SPECULATION SPECULATIVE BUBBLE SWITCH SWITCH SPECULATIVE BUY SYNDACATED LOAN TITOLI DEL TESORO SPECULAZIONE SYNDACATED LOAN TITOLI DI DEBITO TITOLI DI STATO **SPECULAZIONE** SYNDICATION SPESE DI ISTRUTTORIA TITOLI GARANTITI SPEZZATURA TITOLO DI CAPITALE SPIN TAEG TITOLO NON QUOTATO SPIN TITOLO SOTTOSTANTE TAH SPLIT TAKE PROFIT TITOLO SOTTOSTANTE SPOSTAMENTO TAKEOVER TITOLO SPAZZATURA SPOT SPOT CONTRACT TMT TO ROUND TAN TANGENTE SPOT CONTI TANGIBLE FIXED ASSET TOMBSTONE TARGET PRICE TOMBSTONE TOP-DOWN APPROACH TOTAL RESERVES TOTAL YIELD SPOT PRICE TARN TASSA SUL GUADAGNO IN CONTO SPREAD SPREAD CAPITALE TASSI DI BASE TRACKING ERROR SPREAD SPREAD TASSI DI CAMBIO TRACKING ERROR SPREAD A FARFALLA SPREAD A TERMINE TASSI DI INGRESSO TRADE TASSI DI INTERESSI A BREVE TRADE AGREEMENT SPREAD CON OPZIONI TASSI DI INTERESSI FUTURE TRADE BALANCE TASSI DI MERCATO MONETARIO TASSI INTERBANCARI TRADE CREDIT SPREAD DENARO SPREAD DI PREZZO

TRADE DEFICIT TRADER TRADER TRADING CURBS TRADING PHASE TRADING PHASE TRADING RANGE TRADING RANGE TRADING VARIATION TRANFERT AGENT TRANSACTION TRANSACTION ACCOUNT TRANSAZIONE

TRANSFER TRASLATION **TRATTA** TRATTA

TRATTA A SCADENZA FISSA

TRATTA A VISTA TRATTA BANCARIA TREASURIES TREND

TREND TREND LINE TRENDING PHASE TRENDING PHASE TRENDLINE TRENDLINE TREYNOR RATIO'S TRIANGOLO

TRIPLA A
TRIPLE WITCH HOUR TRIPLO MASSIMO

TROUGH TRUST TRUSTEE TURNOVER TUS - BANK RATE
ULTIMO GIORNO DI CONTRATTAZIONE

ULTIMO PREZZO

UNAMORTIZED BOND DISCOUNT UNAMORTIZED PREMIUM ON

INVESTMENT UNCALLED CAPITAL UNCOLLECTED FUNDS UNCOLLECTED FUNDS UNCOMMITTED UNCOVERED OPTION UNDER REVIEW
UNDERCAPITALIZATION UNDERLENDING

UNDERLYING

UNDERLYING FUTURE CONTRACT UNDERLYING FUTURE CONTRACT UNDERLYING INDEX

UNDERWATER LOAN UNDERWATER LOAN UNDERWRITER UNDERWRITING UNDERWRITING UNDERWRITING

UNDERWRITING AGREEMENT UNDERWRITING AGREEMENT UNDERWRITING COMMISSION UNDERWRITING GROUP UNDERWRITING GROUP UNDERWRITING SPREAD UNISSUED STOCK

UNIT

UNIT UNIT

UNIT INVESTMENT TRUST UNIT INVESTMENT TRUST UNIT LINKED

UNIT OF TRADING UNITA' UNITA'

UNITA' DI CONTRATTAZIONE UNITA' DI CONTRATTAZIONE UNIVERSAL BANK

UNLIMITED TAX BOND UNLISTED MARKET UNLISTED MARKET UNLISTED SECURITY

UNLISTED TRADING UNLOADING UNMATCHED BOOK UNPAID BALANCE UNPAID DIVIDEND UNREALIZED PROFIT UNSECURED CREDITOR

UNSECURED DEBT UNSECURED LOAN UNSECURED LOAN UNWINDING LIP

UPGRADING UPSIDE POTENTIAL UPSTAIRS MARKET UPTICK

UPTICK UPTICK RULE **UPTICK RULE** UPTREND

UTILE AL LORDO DELLE IMPOSTE

UTILE DA INVESTIMENTO NETTO PER

AZIONE UTILE LORDO UTILE NETTO

UTILE NETTO PER AZIONE UTILE OPERATIVO UTILE OPERATIVO UTILE PER AZIONE

UTILI ANTE IMPOSTE UTILI NON DISTRIBUITI UTILI PER AZIONE UTILI PER AZIONE DILUITI

LITH ITA UTILITIES UTILITY

UTILIZZO DEL FATTORE LAVORO

VALIDATION VALIDAZIONE VALIDO QUESTO MESE VALORE VALORE AGGIUNTO VALORE ATTUALE VALORE ATTUALE NETTO

VALORE NOMINALE - FACE VALUE

VALORE RESIDUO VALORE TEMPORALE VALORE TEORICO VALUATION RESERVE

VALUE VALUE ADDED VALUE AT RISK

VALUE BASED MANAGEMENT

VALUE COMPENSATED VALUE STOCK VALUTA VALUTA CHIAVE VALUTA COMPENSATA VALUTA DEBOLE VALUTA DI DENOMINAZIONE

VALUTA FORTE VALUTA REGOLATA

VALUTAZIONE VALUTAZIONE AI PREZZI DI MERCATO VALUTAZIONE CON AVVIAMENTO VALUTAZIONE DEI TITOLI

VAR

VARIABILE PROXY

VARIABLE RATE DEMAND NOTE

VALUTAZIONE DEL CREDITO

VARIANCE VARIANZA

VARIATION MARGIN VARIAZIONE DI PREZZO VARIAZIONE DI QUOTA

VARIAZIONE NETTA VELOCITA' DI MONETA VELOCITY OF MONEY VEND

VENDI AL LIBRO VENDITA

VENDITA A TERZI

VENDITA ALLO SCOPERTO VENDITA ALLO SCOPERTO VENDITA CON GARANZIA VENDITA E RILOCAZIONE VENDITA ECCESSIVA VENDITA IMMEDIATA VENDITA IN INCREMENTO VENDITA SCOPERTA VENDITA SOTTOCOSTO VENDITA SPECIALE VENDITE ELUSIVE VENDITORE

VENDITORE VENDITORE DI OPZIONI VENDOR'S LIEN **VERIFICA**

VERIFICATION
VERSAMENTO IN UNICA SOLUZIONE

VIGILANZA BANCARIA VISIBLE SUPPLY VITA MEDIA VOLATILE VOLATILITA VOLATILITA'

VOLATILITA' ANNUALIZZATA VOLATILITA' IMPLICITA VOLATILITA' IMPLICITA VOLATILITA' STORICA

VOLATILITY VOLATILITY VOLUME DELETED

VOLUME DELLE CONTRATTAZIONI

VOLUME NON INDICATO VOLUME QUOTATION VOTING STOCK **VOTING TRUST** WAIVER OF NOTICE WALL STREET WAP WAREHOUSE

WARRANT WARRANT WASTING ASSET WATCH LIST WATCHLIST WEAK CURRENCY WEIGHTED AVERAGE WHAT IF ANALYSIS

WHIPSAW

WHIPSAW WHOLESALE PRICE INDEX

WIRELESS APPLICATION PROTOCOL WITH HOLDING TAX

WORKING CAPITAL WORKOUT

WORKOUT AGREEMENT

WORLD BANK WRITER WRITER WRITING NAKED WRITTEN DOWN VALUE

XENODOLI ARI XENOMERCATO XENOVALUTE YEN YIELD YIELD

YIELD CURVE YIELD MATURITY YIELD SPREAD YIELD TO MATURITY

YUAN ZECCA

Appendix G Principal Component Analysis for Experimet IV

In the contenxt of Experiment V – Wikipedia Articles trustwhortiness – we performed a PCA of articles, using a set of basic articles' features. This information is required by the trust scheme based on similarity and grouping to understand which features to be used to define standard and which features has to be discarded. The features used in the conputation are:

- Number of edits
- Variance of edits length
- Percentage of Reverted edits
- Average Length of editing
- Variance of sections
- Average length of a section
- Number of discussions
- Number of Images

- Length of the article
- Number of Section
- Number of references
- Number of external link
- Number of notes
- Number of edits in the Talk Page
- Presence of frame

The results of the PCA are the following:

| Feature | Component Score (% Total Variance) | Cumulative Score | Impact | |
|---|---------------------------------------|---------------------|-----------|--|
| Variance of sections | 28.65 | 28.65 | Very High | |
| Length of the article | 19.47 | 48.12 | High | |
| Number of references | 11.39 | 59.51 | High | |
| Variance of edits length | 10.38 | 69.89 | High | |
| Percentage of Contributions by Registered Authors | 4.5 | 74.39 | medium | |
| Number of Images | 4.13 | 78.52 | Medium | |
| Number of edits | 3.27 | 81.79 | medium | |
| Number of Section | 3.49 | 85.28 | Medium | |
| Number of external link | 3.1 | 88.38 | Medium | |
| Percentage of Contributions by Anonymous Authors | 2.94 | 91.32 | medium | |
| Average length of a section | 2.85 | 94.17 | Medium | |
| Average Length of editing | 2.75 | 96.92 | medium | |
| Percentage of Reverted edits | 1.42 | 98.34 | low | |
| Number of edits in the Talk Page | 0.55 | 98.89 | Low | |
| Presence of frame | 0.46 | 99.35 | Low | |
| Number of notes | 0.28 | 99.63 | Low | |
| Number of discussions | 0.18 | 99.81 | Low | |

Features with a low impact are discarded. Features with medium impact are still considered in the computation, while a larger importance is given to the number of references, lenght of articles and variance of sections and edits'length