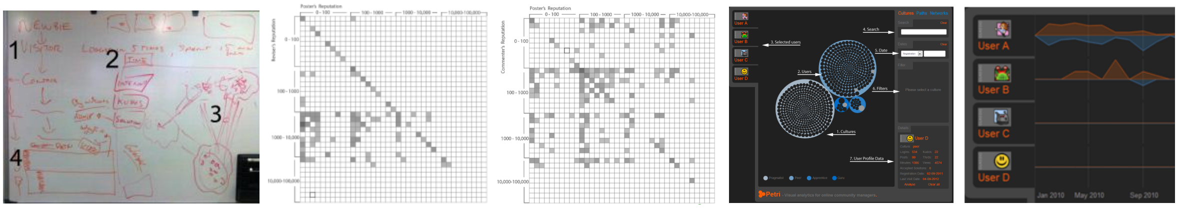
Using visualisation to support Reflective Community Design



A thesis submitted to the

University of Dublin, Trinity College

for the degree of

Doctor of Philosophy

John McAuley

Knowledge and Data Engineering Group (KDEG)

School of Computer Science

Trinity College, Dublin,

John.McAuley@.scss.tcd.ie

Supervised by Dr. Dave Lewis & Dr. Lucy Hederman

Declaration

I, the undersigned, declare that this work has not been previously submitted as an exercise for a degree at this or any other University, and that, unless otherwise stated, it is entirely my own work.

John McAuley

October 2013

Permission to Lend or Copy

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

____

John McAuley

Acknowledgements

Thanks to Dr. Dave Lewis for his support, insight and encouragement. Also, thanks to Dr. Lucy Hederman and Dr. Alex O'Connor for their assistance during the revision process. Thanks to all my colleagues in KDEG and CNGL, both students and staff, who provided invaluable feedback and direction when and when not required. Thanks to both the Super User and Symantec communities, and their respective community management teams, who engaged with the research and participated in the case studies; particularly Jason Rickard and Suzanne Beekes for their interest and time. Thanks to Kevin Koidl, who assisted with the content analysis, pitched in with ideas and helped with review. Finally, a special thanks to my family, supportive as always, and to my wife Síne, fantastic at every turn.

Abstract

Online communities have developed as a mainstay of contemporary internet culture. Their application has enabled the creation of media and software artefacts and resulted in new systems of economic and cultural production. However, online communities are notoriously difficult to build and maintain and their development can often shift in unanticipated directions. This thesis proposes an approach to the development of online communities called reflective community design. The aim of reflective community design is to enable community stakeholders make informed decisions based on an analysis of community behaviour and to learn from these decisions as the community develops over time. At the core of this practice is visual representation. As with traditional design disciplines, such as product design and architecture, visual representation enables designers think about the design space, reason about their design decisions and share their thinking with other designers. Within this context, this thesis introduces a set of design considerations for the visual representation of online communities using techniques drawn from the Information Visualisation and Visual Analytics literature. This framework then guides the development of different visual analytic systems in two independent case studies with different sets of community stakeholders. Both case studies enable the validation of different elements of the framework, allows the author to suggest alternative implementations and propose directions for improvement.

Relevant Publications

Some of the initial ideas discussed in chapters 2 and chapters 3 were presented at:

Dr. Dave Lewis, Kevin Feeney, John McAuley, A Platform for Studying Progressive Self-Management in Online Communities. *WebSci'09 – 1st International Conference on Web Science*, Athens, Greece, 18th–20th March, 2009.

Also parts of this work have been published in the following articles:

The case study in chapter 4 is included in:

John McAuley, Dr. Alex O'Connor, Dr. Dave Lewis, Exploring Reflection in Online Communities. *LAK '12 - Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Vancouver, British Columbia, Canada, 29 April – 2 May 2012, 2012, Pages 102-110.

The case study in chapter 5 is included in:

John McAuley, Dr. Alex O'Connor, Dr. Dave Lewis, Community Management Matters: Advanced visual analytics for online community managers. *Innovative Approaches of Data Visualization and Visual Analytics*. July 2013. Print.

Contents

List of figures

List of Tables

Conventions

In this thesis, we make use of the following conventions:

- We use the term *user* to refer to the user of the visualisation, however user can also denote a community user. Where necessary, we describe community users as community members.
- Conceptually, we consider online communities as a design space and community stakeholders as community designers, however we tend not to use these terms beyond the introduction (chapter 1) and background chapters (chapter 2).
- For convenience, in exposition, *we* refers to the author of this thesis *only*. All of the work in this thesis was conducted by solely by the author.
- For convenience, we use *they* to denote both *male* and *female* gender.

“To see a pattern is often to see a solution to a problem”

Colin Ware, Visual Thinking, 2008

...

“In decision-making the useful information is drawn from the overall relationships of the entire set”

Jacques Bertin, Graphics and Graphic Information Processing, 1981

...

“Society is a joint-stock company in which the members agree for the better”

Ralph Waldo Emerson, Self-reliance, 1841

1

Introduction

> "Communities cannot be declared, but need to be slowly grown over a long period of time."
>
> Aldo de Moor, Using System Dynamics to Construct Design Theory ..., 2007

1.1 Motivation

In October 2010, the following image was sent to the AIR-L (Association of Internet Researchers) mailing list, a large international mailing list devoted to internet related studies:

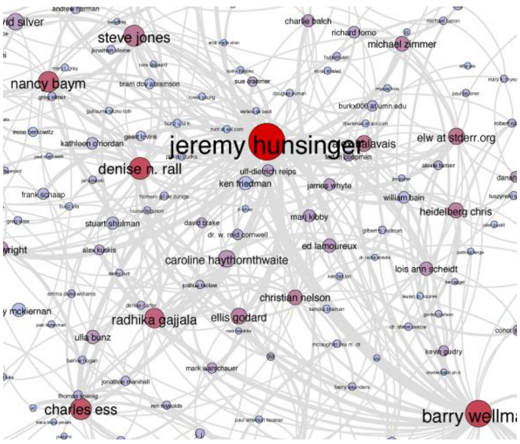


Figure 1: Social Network Visualisation of AIR-L

The image is a social network visualisation of activity on the list over the past number of years. While the visualisation itself is not particularly noteworthy, what is interesting is the reaction that the visualisation provoked. One of the principal contributors replied with the following:

> "Since the policy change on the reply-to[1], I've basically cut my replies by what feels like 90%, so there should be a marked difference between my work on the list from the earlier years and now, and also now I post from two different addresses. In fact, your graph might be an interesting tool to argue that air-l should go back to proper list settings instead of the current ones, but I've given up on that argument..."

This snippet of conversation illustrates how design decisions can be made in online communities. Someone in charge of the technology will, based on repeated requests from sections of the community, make a decision that inadvertently changes aspects of the community's configuration. The extent of this change on the community is difficult to anticipate and is only brought to light in this example with the aid of visualisation[2] technology. The visualisation enabled the community to *reflect* on how this policy change has come to impact the community's flow of communication. This process of reflection was not undertaken privately or by a single individual but was conducted socially and distributed on the community's mailing list. The AIR-L community remains active today, possibly to the point that the decision went unheeded by a majority of users, but the change resulted in a less dynamic and potentially less connected community. In fact, in another circumstance, this decision could quickly silo a community into isolated pockets of connected individuals, making it difficult for the less connected users in the community to contribute towards potentially important discussions. The problem is that, because of the complexity of the community design space and the distributed and asynchronous nature of community interaction, it is impossible for community designers[3] to appreciate how different design decisions come to impact the development of any online community.

As with any human system, online communities generate their own norms and conventions continuously and in accordance to no-predefined pattern or plan. Instead, these structures are contingent on social negotiation and consensus building and while these systems may show periods of regularity, these patterns, as discussed by Truex et al., are often only observable in hindsight (Truex, Baskerville, & Klein, 1999). Thus, communities can never be considered stable or fully formed, but are rather under a continuous process of evolution. Further complexity is evident if we think of online communities in terms of socio-technical systems. The challenge that this presents is not only how to design systems that are fair and equitable, considered by Mumford as a founding principle of socio-technical design (Mumford, 2006), but how to design systems in which both the social and technical elements are considered holistically. A holistic understanding is required because, as illustrated in our opening example, both the social and technical elements are mutually reinforcing - a social change can result in a technical change and a

[1] A mailing list reply-to policy governs whether a reply is sent to the original sender or distributed to the whole list. Often this policy leads to controversy, as the policy dictates whether the list is busy or based on discrete communications.

[2] In this thesis, although we use visualisation as the prominent term, we generally mean information visualisation or the visual representation of abstract concepts and relationships used to amplify cognition and increase human understanding, as originally defined in (Card et al., 1999). At times we also reference visual analytics, which we consider as the use of visualisation techniques coupled with the application of advanced computational techniques to support analytic reasoning (D. Keim, Mansmann, & Thomas, 2010).

[3] Online community designers are considered the class of user most invested in the community's development. Community administrators, list maintainers and community managers are, for the purposes of this thesis, considered examples of online community designers (see 1.1.1).

technical change can equally result in a social change - and anticipating the impact of either is unfeasible due to complex and evolving socio-technical interdependencies (Gerhard Fischer & Herrmann, 2011).

Online community design, thus, is not about designing things, complete in function and form, but is, as discussed by Erickson, about designing things that participate in complex and evolving socio-technical systems (Erickson, 2009). Each design decision, once assimilated into the system, immediately begins to transform the community in some way, however, how this transformation surfaces is difficult to predict. Therefore, what is required is a *design practice* which is based not only on *acting* – or the introduction of design decisions into the community – but also on *reflecting* – or the evaluation of how each design decision comes to impact the community over time. No single design decision should define a community but rather each decision must occur naturally, almost organically, and with the sanctioning of the community, so that when considered together their implementation gradually comes to shape the culture of the community. In this way the community designer is both town mayor, as originally proposed by Preece (Preece, 2000a), and town planner, as proposed by Alexander (Alexander, Ishikawa, & Silverstein, 1977), who, through strategic, continued and informed action gradually nurtures the community's growth. Traditional methodologies with phased implementations and lengthy development cycles are thus replaced with an approach that distributes the design of the system throughout the entire lifecycle of the community[4] (Marty, 2005).

One of the difficulties of considering community design in this way is that, unlike other design disciplines such as product design or architecture, online communities are without a clear representational form that allows designers to systematically reason about their characteristics (Scacchi, 2004). Representation significantly impacts design work, enabling designers to cast their ideas into a material form, which they can then use to think about the design problem and reason about the design space (Hevner, March, Park, & Ram, 2004). This view considers design as a means and end, or a process and an artefact - the process including a set of activities that when committed in sequence produces an artefact. The philosopher Donald Schön describes this process as having "a reflective conversation with the materials of the situation" because the designer, by working on constructs such as sketches and models, represents the design problem and reasons about potential design solutions (Schön, 1992). Of course, representing the complexity of the community design space, given the certainty of change and the distributed nature of community interaction, remains a considerable challenge. Researchers have found that socio-technical diagrams, as proposed by Latour et al. (Latour, Mauguin, & Teil, 1992), and rich pictures, as proposed by Checkland (Monk & Howard, 1998), enable designers and relevant stakeholders to think about complex design problems, however, these approaches are schematic representations that lack resolution and are disconnected from actual practice. Alternatively, as illustrated in the opening example, techniques drawn from Information Visualisation and Visual Analytics provide us with a means to represent complex community dynamics in a meaningful way.

Although a single example, previous research has intimated at the potential for visualisation to catalyse reflection on social behaviour. Warren Sack, for example, describes his Conversation Map - an application that visualises the conversational structures of Usenet - as "discourse diagrams" that provide a set of "conversational foci and a means to reflect upon an ongoing discussion" (Sack, 2000a). Both PostHistory (Viégas, Boyd, Nguyen, Potter, & Donath, 2004) and Themail (Viégas, Golder, & Donath, 2006), two email visualisations developed by Viégas et al., prompted users to reflect on the evolution of their relationships with different people over-time; while Perer and Smith found that visualising email communications between colleagues provoked reflection on the nature and development of

[4] This point is argued more thoroughly in Sect. 2.4.1, however, the principal position is that, due to the number of configurable and extensible web development frameworks and community systems freely available, there is no longer a need for lengthy development cycles. Instead, there is a need to shift the responsibility and awareness of the design process towards individuals invested in the community's development, i.e. community designers.

team dynamics (Perer & Smith, 2006). Similarly, Pousman et al. proposed a casual visualsation, called Imprint, to spark conversation and catalyse reflection as regards paper usage (on printer spools) by groups in large organisations (Pousman, Rouzati, & Stasko, 2008). Although suggestive, this work is hardly comprehensive and there remains a need to conduct further studies in real communities with real community users. Moreover, if visualisation technology can be used to support a more reflective community design practice, which in turn can lead to more robust and sustainable community environments, then it is important that we understand how to design useful and usable visual representations.

1.1.1 Design and Designers

For the purpose of this thesis, *conceptually*, we consider online communities as a "design space" and community stakeholders - such as forum moderators, wiki administrators, community managers and others invested in the community's continued success - as online community designers. This is in keeping with the view proposed by Vande Moere and Purchase of a designer as someone who "devises courses of action aimed at changing existing situations into preferred ones" (Vande Moere & Purchase, 2011) but also recognises O'Day et al.'s view of designers as "insiders to the setting of technology use" (O'Day, Bobrow, & Shirley, 1998). As mentioned, our understanding of the design process is similar to Christopher Alexander's (Alexander et al., 1977) in that we do not suggest that communities are designed, in the way an engineer designs a car. Rather, we consider that each individual design decision, if based on rational analysis and taken with the consent of the community, can incrementally lead to the development of better community environments. By applying techniques drawn from the literature on information visualisation and visual analytics, we propose to render the community design space into visual representations that enable community designers to analyse and reflect on the dynamics of their community.

1.2 Research Question

In this thesis, we address the following research question:

RQ: How can we design visual representations that support reflective community design?

1.3 Research Objectives

To address this research question, we outline the following objectives:

RO1: Establish a basis for reflective community design.

RO2: Analyse how online communities are currently represented using visualisation technology and select a set of design considerations based on this analysis.

RO3: Evaluate the applicability of these design considerations through independent case-based implementations with different users, different communities and different visualisation systems.

1.4 Research Methodology

First we make an argument that cycles of action and reflection, typical of product design, can assist in the development of productive online communities. However, we also argue that unlike product design, online communities lack an accepted representational form. Visualisation, we propose, can be used to represent online communities, providing community designers with a way to reason about community dynamics. This argument is used to motivate the work presented in the remainder of this thesis. Next, we analyse the information and visual analytics literature and synthesise a set of design considerations that can assist in the visual representation of online communities. Unlike visualisation systems designed to support professional analysts, our aim is to design visualisations that support the role of community designer, a user less versed in analytic methodologies and visualisation techniques. Thus, there is a greater need to consider how to address particular aspects of the visualisation design, such as aesthetics, animation and interactivity. Having established a set of design considerations, we set out to evaluate their applicability through a series of case-based implementations. We adopt the case study as a primary research method. Popular in social science research, and increasingly used in information systems and information visualisation research (Shneiderman & Plaisant, 2006), case studies are particularly suited to analysing complex and real-world scenarios in which the phenomenon under observation and its "relation to the environment" are difficult to separate (Flyvbjerg, 2011). This is appropriate for online community research because, as outlined by de Moor, the situated nature of online communities suggests that any particular design solution is only relevant for the instance of that particular design problem (de Moor, 2007a). This is also the case with the application of visualisation technology, what may be considered useful and beneficial in the context of one community may be considered less so in another. We conduct two independent case studies in which we address two community instances with specifically designed visualisation systems. In each case study, we conduct an analysis, select a set of design considerations based on this analysis, implement these design considerations as design elements in a visualisation system and evaluate these design elements in the context of their intended use (Isenberg, Zuk, & Collins, 2008). In following the argument of Ellis and Dix, and the challenge of conducting task-based experiments in ill-defined domains with often incomparable visualisations, we adopt an exploratory methodology for the evaluation of each system (Ellis & Dix, 2006). This approach is also supported by Saraiya et al. who argue that visualsiation is an explorative tool yet a controlled experiement is a confiramative methodlogy, which leaves no room for discovery beyond the scope of a given set of tasks (Saraiya, North, & Duca, 2005). To systematically organise and anlayse the data that we obtain during each evaluation, we use an approach similar to North's insight-based methodology, as described in (North, 2006; Saraiya et al., 2005; J.S. Yi, Kang, Stasko, & Jacko, 2008), and, as proposed by Carpendale, triangulate findings from exit-surveys and exit-interviews (Carpendale, 2008).

1.5 Contributions

The major contribution of this thesis is a set of design considerations that can guide the visual representation of online communities. This approach moves beyond single bespoke implementations towards a unified design framework that can be validated through repeated independent case studies. Given the importance of communities on the web, and the corresponding rise of visual analysis by novice users, there is a need to understand how to design visualisations that can assist in the development of productive community environments. The design considerations presented in this thesis provide visualisation designers and researchers with a generalisable framework to address a broad range of different design problems specific to the visual representation of online communities for online community users.

The minor contribution of this thesis is a demonstration of the applicability of these design considerations through two individual case-based implementations. In particular, we design, develop and evaluate two visual analytic systems that address reflection as individual and collective practice. This contribution is considered noteworthy given the need for more studies in the visualisation literature on how to design systems that address the needs of specific users. Moreover, each implementation enables us to evaluate our design considerations in the "context of their intended use" and to suggest amendments and improvements.

1.6 Thesis Outline

Chapter 2 introduces the background to this thesis. We argue that online communities have complex socio-technical infrastructures and require continuous design cycles in which the community members are community designers. Existing approaches that have lengthy development cycles or rely on external designers are not sufficiently capable of addressing the complexity and continuity required for an online community. Rather, we argue that an approach based on repeated action-reflection cycles, facilitated by visualisation technology, can help address a complex and evolving design space. Action-reflection cycles, it was suggested in the introduction, is a proven design process but the application of visual representation as adjunct to this process in an online community context is less well understood. As a basis for the remainder of this thesis, we suggest there is a need to consider how to visually represent online communities for users that are unversed in analytic methodologies or visualisation techniques.

Chapter 3 investigates how online communities have been visually represented using techniques drawn from the literature of information visualisation and visual analytics. The aim of this chapter is to synthesise a set of design considerations that guide the development of two visualisation systems in the following chapters. In the conclusion to the chapter, the design considerations are summarised and we outline directions for the remainder of this thesis.

Chapter 4 describes a case study in which we apply a subset of the design considerations presented in chapter 3 to the design and implementation of a collaborative visual analytics system called Explore.SU (4.2 and 4.3). The underlying aim of Explore.SU is to catalyse analysis and discussion in an instance of a Q&A community (Super User); however the implementation enables us to evaluate several design considerations in the context of their intended use. We deploy the system on the web, ask members of the Super User community to participate in the study, and observe usage of the system over a two week period (4.4). Based on this approach, we discuss the implications of our design considerations on the Explore.SU system, compare and contrast their applicability and suggest alternative amendments where appropriate (4.6).

Chapter 5 presents a second case study that we conducted with the community management team at Symantec. As with the previous study, this work enables us to implement and evaluate a subset of the design considerations developed in chapter 3. Unlike the previous study, however, this work includes several participative design workshops that enabled us to conduct a domain analysis and select a subset of relevant design considerations (5.2). We describe the implementation of these design considerations in a visual analytics application called Petri (5.3) and evaluate their applicability in several user studies conducted with members of the community management team. Again, we discuss our approach to design, consider the feasibility of the design considerations and suggest amendments in the light of the evaluation (5.6).

Chapter 6 readdresses each of the research objectives outlined in section 1.3 and some potential avenues for further research are identified.

2

Background: Towards Reflective Online Community Design

"Users are the ultimate custodians of and experts in their own practice."

Dan Shapiro, The Limits of Ethnography: Combining Social Sciences for CSCW, 1994

2.1 Introduction

Online communities are an integral part of the web. Once considered as an experiment into the potential of computer-mediated communication, online communities have developed into large-scale systems of economic and cultural production. However, community building remains notably difficult and often significant effort is met with less than significant results. In this chapter, we examine the challenge of online community design by focusing on how to address a dynamic and constantly evolving design space. We analyse several existing methodologies and argue that none sufficiently address the dynamics of online community design. Drawing on the relationship between reflection and action, we propose an approach to augment existing methodologies based on the application of visual representation.

To illustrate how online communities have developed over the last number of decades, Sect. 2.2 discusses the evolution of communities on the web from early social gatherings to contemporary systems of economic and cultural production. Sect. 2.3 examines the challenge of online community design and Sect. 2.4 analyses how existing approaches insufficiently address this challenge. Sect. 2.5 concludes the chapter with a consideration of how reflection, supported by visualisation technology, can address a dynamic community design space.

2.2 The Evolution of Communities on the Web

We consider the evolution of communities on the web in three stages. The first stage involves the development of the first online communities during which entrepreneurs, developers and researchers explored the potential of online social interaction with a variety of novel socio-technical environments. We consider the second stage of communities on the web as integrated with the development of Open Source Software (OSS) and the emergence of the production community, considered by Benkler as groups of self-organised individuals committed to the creation and maintenance of media and software artefacts (Benkler, 2002). Tim O'Reilly's famous essay "What is Web 2.0" helped label many of the developments that emerged during this stage of communities on the web (O'Reilly, 2005). In the most recent stage, communities contribute to almost all aspects of the web, from technical support at large-scale corporations to online learning with the development of the MOOC (Massive Open Online Course). In the rest of this section, we briefly consider each of these stages.

The first stage of communities on the web is intrinsically linked to the development of the internet. The first social interactions were based on the creation of ARPANET and the technical discussions between engineers working on the implementation of the network (Preece, Maloney-krichmar, & Abras, 2003). The emergence of the Bulletin Board Systems (BBS) gave way to communities such as The WELL and the publication of the Rheingold's *The Virtual Community*, arguably the first ethnography of a community on the web (Howard, 1993). Rheingold's descriptions of life in The WELL illustrated the depth of relationship possible in Cyberspace[5] and also brought the idea of the virtual community to the public imagination. People began to turn their attention to how life was developing online. Writing for the Village Voice, the journalist Julian Dibbell described how LambaMOO, the largest Object Orientated Multiuser Dungeon[6], dealt with peculiarly aggressive act by a character called Mr Bungle (Dibbell, 1993). Dibble's article illustrated that, while promising libertarian idealism, the norms that regulate social interaction in the physical world were equally valued online. By examining Usenet through the lens of commons-based governance, Kollock and Smith found that, despite a lack of centralised authority, the community managed to maintain the conversational space in a similar way to how pastoral communities traditionally maintained commons-based resources (Kollock & Smith, 1994). The potential for communities to provide value beyond casual social interaction was also suggested during the first stage of online community development. Hagle and Armstrong, for example, were some of the first authors to argue for a return on community building via user-generated content, customer acquisition (whereby the community provides a market for a particular product) and social marketing campaigns (Hagel & Armstrong, 1997). Similarly Lave and Wenger recognised the potential for technology-supported communities to provide a platform for learning via Legitimate Peripheral Participation[7] (LPP)(Wenger, 1999). This stage of community development also saw the emergence of the first practitioner literature, including Kim's *Community Building on the Web* (Kim, 2000) and Preece's popular *Online Communities: Designing Usability and Supporting Sociability* (Preece, 2000a), both of which provide guidance on the development of productive, mainly small to medium scale, online communities. Both authors recognised that community building is as much a social as it is a technical process and that successful strategies involve an integration of the two.

---

[6] Object Orientated Multiuser Dungeons, often described simply as MUDs or MOOs (MUD, Object Orientated), are text-based real-time virtual environments in which users create objects and spaces using text-only commands (Preece, 2000b).

[7] LPP, originally coined by Lave and Wenger, is the process in which new entrants into a community learn, through repeated interaction with existing members and by undertaking initially peripheral but increasingly central duties, to become part of a community (Lave & Wenger, 1991). LPP has been described as the "Apprenticeship" model of learning and is associated with communities of practice and the process of learning by doing.

We consider the second stage of communities on the web as connected to the development of Open Source Software (OSS). During this stage, Richard Stallman introduced the notion of free software[8] and both Linus Torvalds' Linux operating system and the Apache Software Foundation's Apache web server spearheaded the uptake of OSS technology. Licensing structures, such as Apache, MIT and GPL (General Public License), became an accepted way to share or freely distribute software on the internet (Weber, 2005). In 1999 Eric Raymond wrote *The Cathedral and Bizarre*, a description of the culture and organisation intrinsic to OSS (Raymond, 1999). The Cathedral and the Bizarre was a seminal publication in that it not only popularised the OSS development methodology, and the idea of Linus' law (given enough eyeballs all bugs are shallow), but also illustrated what could be achieved by loosely affiliated communities of geographically dispersed but intrinsically committed individuals. In 2001, Lawrence Lessig's Creative Commons took the basic OSS licensing framework and applied it to the generation and distribution of any media content on the internet (Lessig, 2006). The success of Open Source, or more importantly the licensing framework and organisational structure that underpinned open source, proffered a new stage for communities on the web. Once considered a place of discussion, online communities were now considered a place of production. Unlike OSS communities however, other forms of production communities required users to have no specialist technical skills or expertise. This enabled production communities to broaden their membership base and include users that could never have contributed to OSS projects beyond bug reports and feature requests. Although communities such as Slashdot had shown the potential for peer production, Wikipedia, with its lack of technical or governmental constraint, emerged as the prototype for this new form of production community. The ability for anyone to contribute in whatever way the wanted, by essentially ignoring all rules[9], enabled Wikipedia to scale rapidly without the constraints imposed by similar projects. This was both a social and technical phenomenon as many of the principals of self-organisation on the internet, first examined by Kollock and Smith in their analysis of Usenet, were realised to their full potential with Wikipedia. The ability for the community to not only to develop their own system of governance but also to develop their own means of participation suggested new methods of cooperation originally thought impossible[10]. Naturally, several seminal publications arrived on the back of the community's success, for example, Benkler's *Wealth of Networks* (Benkler, 2006) and Tapscot and William's more accessible *Wikinomics* (Tapscott & Williams, 2006), both arguing for a community-centric approach to production that is founded on participation, transparency, openness and sharing, while Tim O'Reilly published his famous essay "What is Web.20" identified social interaction as a core component of a contemporary web experience (O'Reilly, 2005). While this stage of community development made clear that communities can produce media as well as software artefacts, it also made clear that community design was based on a complex set of evolving socio-technical conditions. The introduction of production required the designers to not only think about designing an active social space but also to think about designing a system that can support a production process.

The latest stage of community development is defined by a continued integration of communities into almost every aspect of the web, coupled with the emerging interest in the application of analytic technology. Commercial enterprises, non-for-profit organisations, political and commercial advertising campaigns, all make use of online

[8] http://www.fsf.org/

[9] This is a reference to Wikipedia's founding principle of "Ignore All Rules" that has since been replaced with a large bureaucratic system of rules (Butler et al., 2008). Increasingly, this has been referenced as one of the reasons for decrease in contributor activity in Wikipedia (Simonite, 2013).

[10] Initially Mancur Olson, who first proposed "The Tragedy of the Commons", argued that group size is inversely related to the cooperation (Olson, 1965). Thus, the larger the group size, the less potential the group has at sharing common interest, and the less potential the group has at acting collectively. Kollock and Smith argued against this view with their analysis of Usenet, similarly the success of Wikipedia has also shown that this principle holds little basis when applied to internet scale communities.

communities to evaluate their products and propagate their message. Business interests have recognised the possibilities of peer-production and sought to help organisations identify how best to approach the creation of collaborative and crowd-sourcing environments (Haythornthwaite, 2008; Pisano & Verganti, 2008). Companies, such as Symantec, Microsoft, Google and Facebook, maintain corporate sponsored online communities to assist in the provision of technical support for both customer and developer networks. MVP, or Most Valued Professional, programs are used to motivate participation rates and encourage a deeper level of commitment (Preece & Shneiderman, 2009). The success of the Khan[11] and Code[12] academies have illustrated the real-world potential of online learning and given rise to the MOOC, or Massive Open Online Course, now a mainstay of many large Universities. Nevertheless, the challenge of community design remains. Retaining community members is increasingly difficult given the sheer number of potential offerings available on the internet. MOOCs, for example, have notoriously low completion rates in comparison to the number of individuals that sign up for any particular course (Parr, 2013). Similarly, in an interview with Gawker, Jimmy Wales, the founder of Wikipedia, has expressed concern as regards the long term viability of the project given the reduction in the number of individuals contributing the encyclopaedia's development (A. Chen, 2011). Thus, while communities continue to offer huge potential, in terms of economic and cultural production, their design and continued proliferation remains a challenging undertaking. Both commercial and research interests have converged in trying to understand how to build and maintain sustainable online communities. In Web Science and Social Computing, for example, researchers are increasingly encouraged to move beyond the lab to field sites and conduct real-world experiments that "capture ecologically valid situations" (Bernstein, Ackerman, Chi, & Miller, 2011; Chi, 2009). Likewise, commercial organisations, such as the Stack Exchange Network, have developed a sort of "living laboratory" (Suh, Convertino, Chi, & Pirolli, 2009) where approaches, such as A/B Testing, enable experimentation with different communities under different conditions. In almost all cases analytics – or the application of computational techniques coupled with the use of visual interfaces to enable the examination of large datasets - is playing a progressively important role. This is borne out by the increasing number of publications that focus on the application of analytic technologies. Much of this work involved researchers analysing social media corpora but more recently novel approach involve increasing the resolution of particular analytic technique. For example, Chan and Hayes use a variety of social network features to decompose the Boards online community into a number of different social roles (Chan & Hayes, 2010), Wagner et al. use a range of machine learning models to analyse attention patterns again using the Boards community dataset (Wagner, Rowe, Strohmaier, & Alani, 2012), while Furtado et al. use a combination of cluster analysis and Artificial Neural Networks (ANN) to consider the contributor profiles of users in the Stack Exchange Network (Furtado & Andrade, 2013). Although this work contributes towards our understanding of the structural organisation of large-scale online communities, there is an increasing need to understand how to apply this sort of research in the process of online community design.

2.3 The challenge of online community design

Although we could consider the challenge of the online community design from a range of different perspectives, such as critical mass and anti-social behaviour, we focus specifically on how community designers address a complex and evolving design space. Online communities are increasingly discussed in terms of socio-technical systems in which change is not only constant but the outcome of change is difficult to anticipate (de Moor, 2007a; Jahnke, 2009). In this section, we consider how complexity and change can challenge the development of productive online communities.

---

[11] www.khanacademy.org

[12] www.codecademy.com

2.3.1 Addressing a Continuous Design Cycle

Addressing constant change suggests the need for a continuous design cycle - or one that can manage a design space that is in constant flux. Socio-technical systems designers do not aim to create a completed system but to continually and gradually improve system behaviour and performance over time. However socio-technical systems are reciprocally related and their development is based on a process of co-evolution. This means a change to the technical system can have an impact on the social system while equally a change to the social system can have an impact on the technical system (Gerhard Fischer & Herrmann, 2011). Furthermore, given the inherent complexity of the system, once a design decision is made, the implications of that decision immediately proceeds to transform the system, often in unexpected ways (O'Day, Bobrow, & Shirley, 1996; O'Day et al., 1998). How design decisions surface in a community, thus, is difficult for a designer to anticipate and their observations as regards the outcome of a decision is often made in retrospect. Research on socio-technical systems design has indicated the need for design to be distributed through-out the entire lifetime of a system and for designers to develop a deep understanding of socio-technical co-evolution (Marty, 2005).

2.3.2 Managing the Process of Co-creation

Although initially, much of the work required to maintain an online community is undertaken by a select group of community members – founders, moderators and administrators for instance – for a community to become self-sustaining there is a need to increasingly include a broader set of community members in the construction of the community's environment. This includes decision-making at various levels but also considers how to enable community members to adapt the community's environment to suit their particular needs. Forte and Bruckman's analysis of Wikipedia, for example, illustrates how sub-communities develop their own sets of norms, polices and general rules of governance (Forte & Bruckman, 2008). Enabling community members to take an active part in the creation of the community environment also reduces the potential for abandonment while increasing the sense of ownership (Gurzick & Lutters, 2009). The ultimate example of co-creation can be found in the constructivist design that emerged in the MUD environments of the late 90's, however, no matter in what capacity co-creation is addressed, there remains a need for community designers to exert some level of control in regards to how social norms and policies are upheld and maintained.

2.3.3 Community Members are Community Designers

It is unlikely whether someone external to the community process has the ability to address a continuous design cycle. Typically, designers are involved at the commencement of a project, when prototyping or during specification. However, as discussed in Sect. 2.3.1, complexity coupled with change requires that design is not conducted at a particular stage of the community's development but is rather distributed throughout the entire lifecycle of the community. O'Day et al. argue that in online communities it is important for "designers to be insiders to the setting of technology use" so that they can address the design problem from an informed perspective and consider both the social and technical elements together (O'Day et al., 1998). There is also the need for legitimacy. Ostrom, having conducted extensive studies on self-governing communities, stipulates that decisions must be taken from within a community and not imposed by some external authority to a community (Ostrom, 2000). Similarly, Whitworth and de Moor argue that community members have rights and that those rights must be respected if a change to the community's governing framework is under review (Whitworth & de Moor, 2002). Legitimacy helps build trust in online communities, which in turn builds social capital and social capital, which is, as considered by Preece, integral to developing sustainable community systems (Preece, 2004).

2.3.4 Thinking about Socio-technical implementation

It is important that designers consider both the social and technical elements of any community system together. Several authors in the literature on socio-technical design describe this as taking a "holistic" view of the system, one that includes an understanding of how the social and technical elements integrate (Barton, Emery, Flood, Selsky, & Wolstenholme, 2004; Goldspink & Kay, 2007). However, technical decisions can have a more significant impact on a community than social decisions. Butler et al., for example, describe how violations of 3-R-R policy in Wikipedia - a policy that states a 24-hour ban to users who make three reverts to a single article in any 24 hours - are not automated but are rather arbitrated by Wikipedia administrators (Butler, Joyce, & Pike, 2008). Implementing this decision socially, as argued by Liu and Dix (Liu & Dix, 1997), enables individuals to exercise discretion in how the decision is administered. However, if this decision is implemented technically, then the user receives an automatic 24-hours ban and the circumstances of the ban are incidental. Community designers do not wish to alienate users who may have misunderstood a community policy or are unaware of a particular convention, yet, there remains a need to maintain the social space for the rest for the community. As result, often, as illustrated with this Wikipedia example, there is a preference for social as opposed to technical decisions because fundamentally technical decisions determine social behaviour. However, as a community scales, there can be a need to implement design decisions at a technical level.

2.3.5 Representing the Online Community Design Space

Finally, there is the question of representation. Hevner et al. argue that "representation has a profound impact on design work" (Hevner et al., 2004). As discussed in the introduction, traditional design disciplines, such as product design and architecture, have accepted representational forms that enable the generation of artefacts as part of a design process. Plans and schematic representations, for instance, allow designers to think about the design space and share their thinking with other designers. As these conversations progress, the artefacts, themselves, are adapted and refined as part of an evolving collaborative design process. Online communities, however, are complex and evolving socio-technical systems, distributed in nature and subject to continuous change. How can we capture and represent this complexity in such a way that allows designers to think and reason about their community? Moreover, given design is a collaborative process, how can we reduce this complexity into simple representations that can be shared with other designers or indeed communicated to relevant stakeholders? Any valid means of representation must not only address the design space but also consider the users of the representation. Monk and Howard argue that Rich Pictures - the simple diagramming tool that emerged from Peter Chekland's soft-systems methodology - enables designers to include different stakeholders in the design process without burdening the user with complex visual aids (Monk & Howard, 1998). While it is likely that the simplicity of the approach increases the size of the potential audience, the ability of simple conceptual representations to capture the dynamics of evolving community systems remains questionable. In contrast, visualisation, as developed in the fields of Information Visualisation (Card, Mackinlay, & Shneiderman, 1999) and Visual Analytics (Thomas & Cook, 2005), can, as illustrated in the opening example of chapter 1, represent complex community dynamics yet engage novice users. For example, Viégas and Wattenberg's work on Wikipedia illustrates the potential for visualisation to capture the dynamics of large community systems (Viégas, Wattenberg, Kriss, & Ham, 2007; Wattenberg, Viégas, & Hollenbach, 2007), while their work on Many Eyes shows that visualisation can engage new audiences, generate discussion and enable social data analysis – an approach to analytics performed collaboratively by non-expert users (Jeffrey Heer & Agrawala, 2008; Viégas, Wattenberg, van Ham, Kriss, & McKeon, 2007). There are, however, additional concerns. Bias can creep into the community design process, particularly if, as discussed by Levitt and March, design decisions are made without some degree of objective analysis (Levitt & March, 1988). Online community design involves encoding inferences from the community's history into routines that come to determine or influence community behaviour. Often, those closest to the phenomenon under observation do not make the most objective observers, which can, in turn, result in disproportionate levels of action in response to the interpretation of events. A valid approach to representation, however, can help circumvent many of these concerns, providing community designers with a way to analyse and reflect on the dynamics of their community.

2.4 Limitation of Existing Methodologies

Having considered the challenge of online community design from the perspective of complexity and change, this section examines the limitations of existing methodologies. To do this, we first address how information systems are traditionally developed – a staggered or staged model of development that focuses on the design and implementation of bespoke software solutions (Sect. 2.4.1). Second, given the emphasis on co-creation, we consider participative/participatory design – a set of design methodologies that attempt to embed the user into the design process (Sect. 2.4.2). Finally, we consider community design methodologies – a range of strategies that focus on participation and socio-technical implementation (Sect. 2.4.3).

2.4.1 Traditional Development Methodologies

Traditionally, systems development is conducted in a number of discrete stages. de Moor, in a broad analysis, divides these stages between analysis and design (de Moor, 2007a). System analysis involves the rigorous and systematic organisation and examination of data about the system and the environment of its intended use. System design involves the generation and formation of a new system based on this analysis. Bridging the two stages are a set of requirements and a specification document. The requirements represent what the user needs from the system while the specification document represents a formal description of those needs as understood by the system developer. In most cases, this is described as requirements engineering. The intention behind the staged implementation is to rationalise the process and ensure that the resulting software artefact addresses user requirements. Typical methodologies, such as Waterfall, provide the necessary structures and mechanisms to take a set of requirements and produce a software artefact. Realising the difficulty of incorporating new requirements into a sequential development lifecycle led to the introduction of iterative development methodologies (such as Spiral and later Agile) with incremental and customer focused releases (Larman & Basili, 2003). Nevertheless, as de Moor argues, in both cases the development lifecycle is discrete in the sense that new versions with increasing functionality are developed and then released in discrete intervals (de Moor, 1999). At some point in time, often due to budgetary constraints or some other external influence, the development cycle completes and the system is maintained but the functionality is never extended. This approach has proven successful at addressing end-user development whereby users are primarily interacting with the system but not with one another. However, there are a number of shortcomings when we think about online community design and the application of this methodology. First, as discussed in Sect. 2.3.1, because communities are contingent on social interaction and social interaction leads to unpredictable system requirements, communities require a continuous development methodology as opposed to a discrete one. A continuous development methodology allows for the implementation of design decisions gradually, almost organically, and in line with the community's ongoing social development. This approach reduces the potential for the community's technical platform to fall out-of-sync with the community's social requirements, a situation described by Ackerman as the socio-technical gap (Ackerman, 2000). Second, traditional systems development is based around producing technology to address a particular set of requirements, however, as discussed in Sect. 2.3.4, community design requires the introduction of socio-technical design decisions to address socio-technical design problems. Socially acceptable design decisions, such as policy amendments, can be preferable to rigid technical implementation due to the constraints that technical implementation imposes on social interaction. Finally, the advances in community software such as the plug-and-play open source platforms now freely available, has in many cases reduced the need for discrete and lengthy development cycles. This has removed much of the technical challenge and allowed community designers to focus on building sustainable community systems.

2.4.2 Participative Design Methodologies

The difficulty of drafting a specification that accurately reflects user requirements illustrated a need for a more active involvement from users. Several design methodologies, ranging from participative/participatory design to cooperative design, emerged to meet this need. In each case, the aim is to increase the level of active user participation. Participative/participatory design, for example, involves the user working actively with the system designer to explore the affordances and constraints of a specific technology within a specific context. The system designer, playing the role of facilitator, attempts to draw the user into the design process. Cooperative design argues that both the system designer and user of the system should cooperate towards the goal of a completed artefact, a process in which both parties learn from each other (Bødker, Grønbæk, & Kyng, 1995). Although, each approach proposes a variation on the theme, the underlying principles remain the same, the user learns about the possibilities of the technology while the system designer learns about the context of use. Of course, in terms of large scale organisational investment, participatory design approaches yield significant benefits, not only in terms of requirements gathering but also when addressing political difficulties, between management and staff for example, and improving rates of adoption (Kensing, Simonsen, & Bodker, 1996). However, despite the focus on participation, a number of inconsistencies remain when we consider the challenge of online community design. Most notable is the reliance on a systems designer to facilitate the design process. This is problematic for a number of reasons. First, typically participative design sessions are undertaken at the commencement of a project so that the system designer comes to appreciate how a particular technology will address a particular set of problems given a particular social context. However, because, as discussed in Sect. 2.3.1, it is impossible to anticipate how different design decisions will come to impact a community, online communities require a continuous design cycle or one that is distributed throughout the lifecycle of the community. Second, relying on an external designer will, regardless of how well coupled this process is to the community's existing design practice, result in a discrete development cycle as the system designer's decisions are incrementally integrated into the community's development cycle. Third, participative design is focused on providing a technical solution to address a socio-technical design problem. However, as realised in Sect. 2.3.4, social decisions in online communities can be used to circumvent technical implementation. Finally, there is the issue of resources and the difficulties that this can present communities if they wish to engage external personnel. However, given the importance of co-creation, a range of approaches have been implemented that allow for participation but remove the need for an external designer. Feedback widgets, such as User voice[13], User echo[14] and Get satisfaction[15], provide a means for community members to contribute to the continued development of their community without the cost or other overheads associated with traditional participative design.

[13] www.uservoice.com

[14] userecho.com

[15] getsatisfaction.com

2.4.3 Community Design Methodologies

Several additional community-orientated methodologies have been suggested to address the limitations of traditional development and participative design methodologies. One of the most notable is de Moor's legitimate user-driven specification method, a formal approach to the specification of community systems that builds on canonical theories such as Habermas' theory of communicative action and Searle's speech/act theory. The aim of this approach is to remove the designer from the specification process by providing a formal specification tool that enables a rational conversation as regards the specification of community systems (de Moor, 1999). However, although the schema provides a general way to conceive of specification in a coordinated yet rational fashion, it is complex and restrictive in the sense that participants in a "conversation for specification" must think about the nature of their communication before committing to it. This criticism has been levelled against formal conversational models more generally. Suchman, for example, argues that communication is situated and that meaning or intention cannot be categorised prior to the actual interaction (Suchman, 1993) while Ljungberg and Holm argue that formal conversational models lack the flexibility required to address many social contexts (Ljungberg & Holm, 1996). Nevertheless, communities have developed conversational spaces to assist in the process of design, but these conversational spaces have been modelled on their existing community environments and not based upon rational models of conversation. The most notable approaches include Wikipedia's Village Pump and Stack Exchange's Meta-sites. Preece's Participatory Community-Centred Development (PCCD), in contrast, provides high-level and phased guidance when building online communities (Preece, 2001). While emphasising the interplay between the social and technical design elements, promoting, as the author describes it, "community solutions and technical solutions", PCCD focuses on soliciting continuous feedback from users. While PCCD and variations on the theme remain popular, the approach neglects to account for how visual representation - or the application of analytic technologies - could be used to support the design process. Alternatively, Iriberri and Leroy propose a set of design principles to address different stages of the online community lifecycle (Iriberri & Leroy, 2009). Although static design principles are useful, particularly when confronted with a recognised design problem, their application can never sufficiently address a dynamic design space. As a result, CommuntiyCompare (Xu, Chen, Matthews, Muller, & Badenes, 2013) provides community leaders with simple visual widgets to assist in the analysis of their online communities, however, the approach only addresses the comparative analysis of corporate sponsored online communities and the categories could be broadened to address different communities with different visualisation designs. Finally, Meta-design, by Fischer et al., is an approach to end-user development, which emphasises "under-design", and argues for evolutionary growth driven by seeding and reseeding (G. Fischer, Giaccardi, Ye, Sutcliffe, & Mehandjiev, 2004; Gerhard Fischer & Herrmann, 2011). As with PCCD, meta-design encapsulates many of the principles of online community design as outlined in Sect. 2.3 - including a continuous design cycle and co-creation - however the authors suggest the use of conceptual over analytic visual representations. While, as argued in Sect. 2.3.5, UML or Gantt charts can assist stakeholders to think the problem domain; each approach is a conceptual representation that lacks resolution and is disconnected from actual practice.

2.5 Reflective Community Design

Although a range of methodologies exist, none are sufficiently capable of addressing the challenge of online community design. Traditional development methodologies do not support a continuous design cycle, facilitate a design process conducted by community members or propose a valid means of representation[16]. Participative design

[16] The application of schematic representations, such sequence charts and UML diagrams, is primarily restricted to system development.

enables co-creation but fails to support a continuous design cycle, primarily because the systems designer is external to the community process. Community design, in contrast, supports socio-technical co-evolution and the design process is generally based on feedback drawn from the community. However, there lacks a coherent approach to visual representation. This argument is summarised in Table 1.

Table 1: A summary of the limitations of existing methodologies

Design Challenge	Traditional Development	Participative Design	Community Design
Continuous Design Cycle	NO	NO	YES
Co-creation	NO	YES	YES
Designers are Members	NO	NO	YES
Socio-tech Implementation	NO	NO	YES
Valid Representation	NO	NO	NO

The approach that we investigate in this thesis makes use of visual representation to support a more reflective community design practice. Drawing on the writings of Argyris and Schön, reflection provides us with a conceptual framework with which to address the process of change in complex and dynamic situations. This conceptual framework can be applied to different design problems at different levels of resolution. In product design, as discussed in Sect. 2.3.5 for example, designers typically use reflection, assisted by visual constructs and graphical models, to systematically reason about the design space. This is a "trial-and-error" process in which the designer experiments with ideas in action and the consequences of these actions result in further cycles of reflection. This process has been extrapolated to address the design of much larger and more complex systems. In organisational learning, for example, reflection is used to not only "test ideas in action" but also to re-evaluate some of the core assumptions that underpin particular organisational strategies (Høyrup, 2004). Argyris and Schön argue that reflection, when applied in organisational settings, can be used to address the correction and detection of error (described as single-loop learning) but also help to question the variables that govern a particular organisational process (described as double-loop learning) (Argyris, 1976). In each of these examples, however, there is a need for representation. Whether this involves the use of sketches or diagrams in product design or the application of balance score cards in an organisational setting, representation is not only a process of feedback but also enables individuals and groups to systematically reason about different actions and consider potential strategies. The approach we investigate in this thesis makes use of techniques drawn from the Information Visualisation and Visual Analytics literature to represent the behaviour of online community systems. Previous research has shown that visualisation is particularly adept at assisting humans make sense of large amounts of data. Mazza, for instance, describes how our perceptual capabilities enable us to single out points of interest and to identify relationships very naturally (Ricardo Mazza, 2009). Similarly, Munzner argues that visualisation acts as a sort of "external memory" that enables us to develop internal representations or "mental maps" of specific information spaces (Munzner, 2009a). In terms of socio-technical design, Scacchi suggests that visualisation can be used to cast socio-technical systems into a "representational form" that can then enable designers to systematically reason about their characteristics and dynamics (Scacchi, 2004). Similarly, researchers, such as Víegas, Wattenberg and Heer, have demonstrated the potential for visualisation to support the analysis of data by novice users with little experience of analytic methodologies or visualisation techniques (Jeffrey Heer, Ham, Carpendale, Weaver, & P, 2008; Viégas, Wattenberg, van Ham, et al., 2007). However, there remains a myriad of ways to represent online communities using techniques drawn from this literature and effective representation is not just concerned with the graphical presentation of data but also with the visual competency of the users and the way in which the visual representation will be used in practice. Moreover, considering the situated nature of online communities, an approach to representation that may appear beneficial in the context of one community may have less relevance in the context of another. Thus, as opposed to focusing on the development of a single visualisation system, and evaluating that in the context of one community, we review the literature on Information Visualisation and Visual Analytics and propose a set of design considerations that can be used to address a wide range of different design problems.

2.6 Summary

This chapter looked at the evolution of communities on the web, from early social experiments such as THE WELL to contemporary systems of economic and cultural production such as Wikipedia. Next, focusing on online community design, we examined the challenge of addressing a complex and evolving design space. Existing methodologies are not sufficiently capable of addressing this challenge, given the emphasis on technical over social development, the reliance on external designers or the use static design principles. Conversely, we argued that a more reflective design practice, based on cycles of acting and reflecting and supported by visual representation, can potentially result in more robust community environments. However, given that online communities are situated and context dependent, a representation considered useful in the context of one community, may be considered less so in the context of another. Thus, as opposed to focusing on the development of a system, what is required is a visualisation design framework that can address a broad range of different design problems. In the next chapter we develop this framework by reviewing how online communities are typically represented using techniques drawn from the information visualisation and visual analytics literature.

3

Design Considerations for the Visual Representation of Online Communities

"Graphic representation constitutes one of the basic sign-systems conceived by the human mind for the purposes of storing, understanding, and communicating essential information."

Jacques Bertin, Semiology of Graphics, 1983

"Solving a problem simply means representing it so as to make the solution transparent."

Herbert Simon, The Sciences of the Artificial, 1996

3.1 Introduction

The previous chapter illustrated the complexity of online community design by considering communities as complex socio-technical systems subject to continuous change. It was argued that existing methodologies are not sufficiently capable of addressing the dynamics of online community design and that cultivating a more reflective design practice is required. Although we suggested an approach based on the use of visualisation to support cycles of action and reflection, we have yet to consider how to visually represent online communities. Representing online communities with visualisation technology requires designers to ask themselves a number of important design questions. For example, how can I represent users in a large community? What is the most effective way to address a dense community network? How do I support novice or more casual users? How can I integrate several different visualisations into interactive application? In each instance, the designer must consider the user and create a visualisation that is insightful while also engaging.

In this chapter, we develop a set of design considerations that guide the visual representation of online communities. To do so, we analyse how online communities have been represented to date, using techniques drawn from the Information Visualisation and Visual Analytics literature and, drawing from this analysis, propose a set of considerations that can to address a broad range of different design problems. Effective representation is not just based on the graphical presentation of community data but also considers the users - their visual literacy skills and understanding of the analytic process - and addresses how the system will be used in practice. A set of design considerations, thus, can be regarded as a design framework or intellectual tool that can help to reduce the complexity of the visualisation design process and improve the visual representation of online communities. Furthermore, in later chapters, by evaluating different systems, designed according to this framework, we can also evaluate individual considerations and propose amendments or improvements. Finally, given that the visual representation of online communities has not been previously addressed in this way, we argue that this set of design considerations can be used by other visualisation designers and researchers when addressing online communities with visualisation technology.

First, in Sect. 3.2, we discuss analytic abstraction, often the first stage in information visualisation or visual analytics and considered in this context as a way to provide users with an overview of an online community. In Sect. 3.3, we consider how aesthetics have been applied in the visual representation of online communities. Aesthetics was recognised by Chen in 2005 and by Aslak Burkhard et al. in 2007 as one of the unsolved problems of information visualisation yet remains a challenge for visualisation researchers today (Aslak Burkhard et al., 2007; C. Chen, 2005). Although the aim of information visualisation is to amplify cognition and generate insight, given the burgeoning interest in visualisation by, considered by Pousman et al., a casual audience, visually engaging designs are of increasing importance (Pousman, Stasko, & Mateas, 2007). This is of particular relevance for this thesis given our focus on users with little experience in analytic methodologies or visualisation techniques. We consider a range of approaches from social visualisation - or visualising social information for social purposes - to artistic visualisation - or the use of computational techniques to generate graphic renderings. In Sect. 3.4, we discuss community networks. Social network analysis coupled with node-link diagrams have become a popular way to represent online community datasets; however there are a number of salient points that designers must consider when addressing communities in this way. We address several of these, from spatial layout to the use of matrices to represent dense networks. In Sect. 3.5, we consider the role of interaction and animation. Although static images have expressive value, Yi et al. argue that their utility is limited when addressing larger datasets (Ji Soo Yi, Kang, Stasko, & Jacko, 2007). Interaction and animation can be used to change or enhance a user's mental map while also providing a more engaging user experience. Under the heading Integration and Configuration, Sect. 3.6, we discuss the variety of ways in which different visualisations can be integrated into an end-user application. Again, in this section, we address concerns that the designer may have when developing for novice users. Finally in Sect. 3.7, we consider collaboration, given the emphasis on co-creation in online community design (2.3.2), and analyse different ways in which asynchronous collaborative mechanisms can be integrated with visual analytic systems. In the conclusion to this chapter we summarise each design consideration, identify how and in what capacity the design consideration has been evaluated and outline future directions for the remainder of this thesis.

3.2 Analytic Abstraction

Chi describes analytical abstraction as the process by which structure is drawn from a dataset or corpus of content before visual representation (Chi, 2000). This is particularly important when addressing a large dataset or if a design goal of the system is to provide the user with an "overview" of a dataset (Hornbæk & Hertzum, 2011). The notion of overview was introduced with Schneiderman's information seeking mantra - overview first, zoom and filters then details on demand - but has since been incorporated into the lexicon of information visualisation (Shneiderman, 1996). Overview, in regards to the visual representation of online communities, provides the user with an immediate appreciation of the size and extent of the community, how users in the community relate to one another and, potentially, what kinds of users are in the community. In essence, the aim is to increase the user's awareness of the community, preferably rapidly and pre-attentively. A typical way to provide users with an overview of an online community is through the use of a multivariate scatter-plot, as illustrated by Viégas and Smith's Newsgroup Crowds (Sect. 3.3.3), however there are other approaches that require computation before extraction. Two such approaches, both of which have been applied to online community datasets, include the automatic extraction of both networks and clusters.

3.2.1 Network Extraction

There are a variety of ways to extract a network from a community dataset. The most common is the reply-to network. In a reply-to network, each reply in the conversation is viewed as a link in the network and the authors are considered nodes. The link can be directed, in that a reply suggests an author addressing a second author, or undirected, in which each reply suggests a connection between two authors (Scott, 2000). The difficulty of using a reply-to network is that each approach - both directed reply-to and undirected reply-to - neglects the pragmatics of the conversation. Delugach describes pragmatics as who is talking with whom and in what context is the communication taking place (Delugach, 2007). Reply-to networks are based on the assumption that a reply has significance; however the degree of significance can be difficult to ascertain. For example, a user asks a question in a post to a forum. The answer however, may not be directed to the user but contributed as a reply to another post in the thread. Using an automated reply-to extraction method will fail to register the significance of this interaction and either attribute each reply with the same significance (as undirected connections) or attribute the replies to the user as more significant (as directed connections). In this sense, reply-to networks have a certain naivety as their construction lacks the precision required to accurately model many online conversations.

As an attempt to address these concerns, Gruzd and Haythornthwaite implement an approach in a tool called the ICTA, or Internet Community Text Analyser, which couples a "named network" with a "semantic network" (Gruzd & Haythornthwaite, 2008; Haythornthwaite & Gruzd, 2007). Using entity-name recognition (coupled with an alias resolution mechanism), each user's name is extracted from the body of a message. If the name is at the beginning of the message then the user is considered an addressee. If the name is at the bottom of the message (as the signature) then the user is considered the Original Poster (OP). To calculate the tie strength (or weight) between both the poster and the addressee, the authors measure the number of descriptive concepts[17] that are communicated in a message against all the descriptive concepts in all the messages. This, they argue, gives an indication of tie strength. As implemented previously with Sack's Conversation Map (Sack, 2000a, 2000b), one of the first online community network visualisation tools discussed in Sect. 3.6.4, the semantic network is based on the association of concepts extracted using noun-phrase analysis from the threaded discussions. Although this approach helps to address concerns as regards the pragmatics of the conversation, there are inconsistencies in generating networks using this method. It is difficult, if not impossible, to ensure that all the names are correctly identified and extracted. Changes in relationships can result in different names or "handles" being used by different users at different times. While the "@" symbol has come to represent messages that address specific users (Huberman, Romero, & Wu, 2008), its usage remains restricted to certain social media environments, such a Twitter or Stack Exchange, although it is becoming more universally accepted. Similarly, "concept drift", the process whereby the value and meaning of concepts change over time, challenge the efficacy of automated natural language techniques (Schlimmer & Granger, 1986), particularly when applied to social media environments using open corpus data. As a result, there is no ideal way to generate a network from an online community and each approach will have a degree of inconsistency, however, there are

[17] Their approach to assessing a "descriptive concept" is to extract and measure the number of noun-phrases contained in each message.

several approaches available to the visualisation designer. These observations imply the following design consideration:

DC1. **Network Extraction:** Extract community networks as, either or a combination of, reply-to, semantic or named graphs; consider graphs as directed or undirected.

3.2.2 Clustering

Clustering, in contrast, is the use of computational techniques (described in the literature on machine learning as unsupervised classification) to automatically group users according to a set off user attributes (described as feature vectors)(Flynn, 2000). The application of clustering is far ranging, however, in terms of online communities clustering has been used to organise users into identifiable groupings often described as social roles. Golder & Donath consider social roles as informal social positions taken up by users in a community and are usually based upon "repeated interactions and mutually agreed upon practices" (Golder & Donath, 2004). Familiar social roles include lurker (someone who does not actively engage with the community) and troll (someone who baits other users into arguments). Initially research on the identification of social roles looked towards visualisation as a means to enable the classification of users according to repeated patterns of interaction. Víegas and Smith's early classification, for example, includes debaters, answer people, bursty contributors and new comers (Viégas & Smith, 2004). However, there are difficulties with this approach if we are to consider the size of community datasets and the fluid nature of role composition. To circumvent these concerns, researchers sought to automatically catalogue social roles using unsupervised classification. Chan and Hayes, for example, decompose a large bulletin board community, called Boards.ie, into a number of social roles using a combination of principal component analysis and hierarchical clustering (Chan & Hayes, 2010). One design consideration that we can draw from this work is the selection of social network measures as feature vectors. The authors use a combination of different measures, such as in-degree and out-degree exponents to establish a set of roles that include taciturn, elitist and grunt. This catalogue is by no means definitive but the approach illustrates how relatively simple techniques can be used to develop a useful analytic abstraction. Alternatively, a second design consideration that we can draw from this approach is the use of clustering to develop a set of categories based on user contribution. Preece and Shneiderman argue the online community managers need to better understand the motivational arc of their users so that they can reduce churn rates and increase levels of commitment (Preece & Shneiderman, 2009). Through the application of clustering techniques, we can develop an analytic abstraction that categorises users based on their level of contribution (as shown (Furtado & Andrade, 2013)). This abstraction can then be used by community managers to develop strategies particular to specific user categorises and then to evaluate the effectiveness of those strategies (as users progress along this linear trajectory) over time. This approach favours a feature selection that indicates contribution over structural positioning as suggested by Chan and Hayes. From this analysis, we draw the following two design considerations:

DC2. **Cluster on Interaction**: Cluster users based on repeated patterns of interaction; consider social network measures as feature vectors.

DC3. **Cluster on Contribution:** Partition the community based on level of contribution.

3.2.3 Discussion

In this section we considered how computational techniques can be used to develop analytic abstractions from online community datasets. Analytic abstraction, as suggested by Chi, is one of the first stages of visualisation, prior to visual mapping, and if addressed appropriately can contribute to providing users with an effective overview of the community's social space. Given the inherent challenge of developing networks that represent the pragmatics of online conversation, we considered how to extract both semantic and named graphs. Similarly, we looked at how unsupervised classification can be used to develop role-based categorisations based on structural positioning and contribution rates. In Table 2, we outline how each approach has been evaluated.

Table 2: Evaluation of selected design considerations for Analytic Abstraction

	ICTA	Chan & Hayes	Furtado & Andrade
Network Extraction	**LE**	Ne	Ne
Cluster on Contribution	Ne	Ne	**P**
Cluster on Interaction	Ne	**P**	Ne

P - Prototype only, demonstrated but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not evaluated

Both Furtado and Andrade and Chan and Hayes' approaches to clustering is demonstrated on a number of different communities as opposed to evaluated under controlled conditions; however, given the subjective nature of social role discovery, it is difficult to evaluate this technique using a comparative approach. In contrast, the approach to network extraction explored in the ICTA is compared with existing techniques, namely reply-to networks and existing name extraction toolkits. Their approach performs favourably under laboratory conditions, returning fewer false positives; however generalising beyond this study is problematic given the inherent difficulties of open corpus evaluations with social media content.

3.3 Aesthetic Visual Mappings

Aesthetics can sometimes be overlooked in information visualisation, given the tendency to favour function over form, however the increasing interest in the analysis of data by novice users has helped to emphasise the importance of creating aesthetically pleasing visual representations. To engage new audiences with information visualisation, designers need to create visual designs that are not only insightful but also intuitive and engaging. Aesthetics, from this perspective, can influence each aspect of a design process, from how the visual variables are initially encoded, to how a graph is spatially positioned or interactivity implemented. However in each instance, the role of aesthetics, as argued by Lau and Vande Moere, is not to replace but to augment function (Lau & Vande Moere, 2007).

In this section, we address visual mapping - or how each variable is visually encoded on a 2D plane - and focus on a sub-category of information visualisation called social visualisation. Social visualisations are visual descriptions of social spaces that make social interactions, and other forms of important social information, visually salient (Karahalios & Viégas, 2006). The use of the term "description" is important because the aim of a social visualisation is often not to create an accurate representation but to convey the meaning - the sense and substance - of a social space. Although researchers are investigating the use of social visualisation in public spaces, notably (Vande Moere & Hill, 2012), the majority of approaches have involved the visualisation representation of online communities.

3.3.1 Playful Visual Mappings

Although traditionally the aim of information visualisation is to convey data in as an objective a means as possible, existing research has proposed a range of alternative visual mappings based on playful visual metaphors.

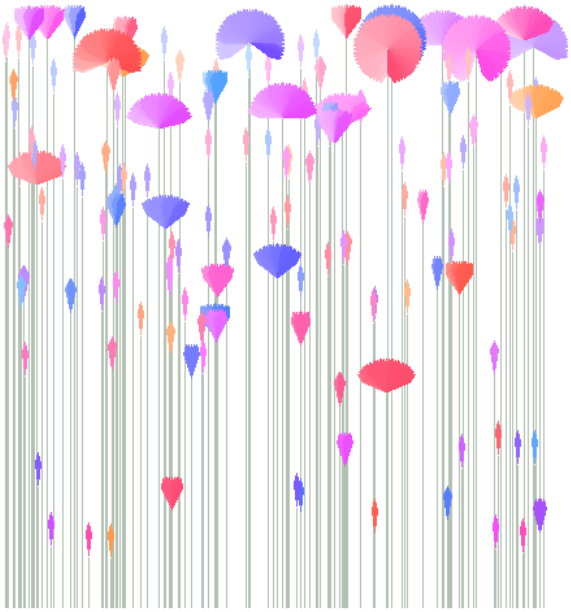


Figure 2: An example of a People Garden visualisation (Donath, Dragulescu, Zinman, Viégas, & Xiong, 2010; Xiong & Donath, 1999). **Each user is represented by a flower. The stalk of the flower represents the length of time that a user has been active in the community. The radial scope of the flower, comprised of petals (each representing a post), illustrates the length of time since a user's last post. The colour identifies whether it was an initial post or a reply. Replies to posts are represented by small dots trailing from the petal, not visible in the current image. A user, when addressing People Garden, can see if a community is active and "blooming" or if a community is "inactive" and withering while also considering how each individual user contributed to the community.**

In People Garden (Figure 2), for example, Xiong & Donath propose the notion of a "data portrait" - a novel visual description of a user in a community based on their history in the community (Donath et al., 2010; Xiong & Donath, 1999). One of the principles of a data portrait is not to realistically or accurately portray the user but to visually depict the user in some abstract way. Often, this can involve the application of a playful visual metaphor. In People Garden, users are represented as flowers and the community is collectively rendered as a garden. Retinal variables[18], such as the colour of petals and the length of each flower's stalk, are used to associate users with similar patterns of posting behaviour. One design implication that we can draw from this approach to is the ability for organic metaphors to evoke our understanding of fundamental social processes and for playful visual mappings to pique our interest and provide an engaging experience. Using People Garden, for example, "newbies" or new community members can quickly identify with the community - active communities bloom while inactive communities wither - but also identify the degree to which each user contributes to the community. The approach is interesting and accessible to novice visualisation users. However, addressing a large community with several thousand users will compromise the integrity of the visual representation, increasing the potential for the display to become dense, cluttered and difficult to read[19]. A second difficulty lies in the author's approach to "data portraiture". A data portrait,

---

[18] Retinal variables, as originally defined by Bertin, are readily distinguishable visual encodings such as size, shape, texture, colour and orientation (Bertin, 1983c).

[19] Although research has shown that pan and zoom can facilitate analysis (Ji Soo Yi et al., 2007), salient points need to be identifiable at a macro level to warrant further investigation at a micro level. Extrapolating the data into higher order renderings, such as

as described by Donath and Dragulescu, is a specific approach to social visualisation that attempts to convey some aspect of a user's character or position in a community (Donath & Dragulescu, 2009). Unlike traditional Information Visualisation, which strives towards objective representation, the designer of a data portrait creates a subjective rendering that mediates the viewer's interpretation of the underlying data. Stressing this point Donath et al. add that "the artist's personal style and interpretation are important aspects" of a data portrait (Donath et al., 2010). By ignoring convention, however, the design of a data portrait is open to be read and interpreted in a range of different ways[20]. Ware argues, in *Information Visualisation Perception for Design,* that arbitrary and unconventional visual representations are "hard to learn", "easy to forget" and "prone to cultural interpretation" (Ware, 1999), while Smuc et al. have shown that reading novel visualisation remains challenging, even for experienced users (Smuc, Mayr, & Lammarsch, 2008). Finally, although we argue that playful visual mappings can provide an engaging experience, particularly when addressing novice visualisation users, People Garden is never evaluated, which would not only benefit the work but would lend support to the argument. From this perspective, there is scope to conduct more extensive research on the practical benefits of playful visual mappings.

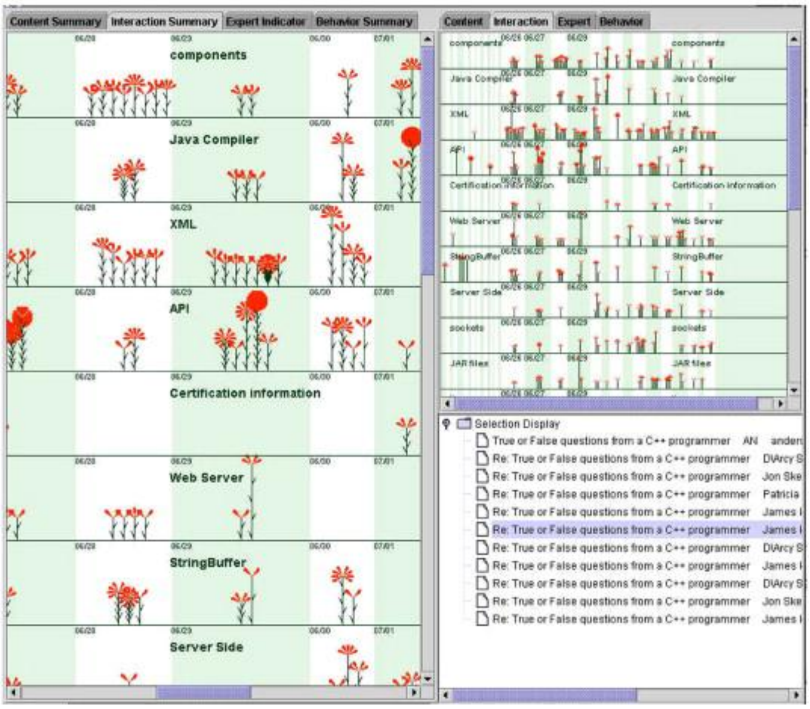


Figure 3: Communication Garden is a visualisation of an organisational communication archive (Zhu & Chen, 2008; Zhu, 2002). Users and threads are visualised as flowers.

A second visualisation that, again, draws on an organic metaphor is Communication Garden by Zhu and Chen (Zhu & Chen, 2008; Zhu, 2002). The authors argue that Communication Garden provides a way for new users to understand how to participate and integrate into an online community. The focus is not solely on analysis, therefore, but also on

---

clusters in a "spatialization" diagram, as described by Skupin for example (Skupin, 1996), is not possible using the current visual metaphor, although, the data can be re-encoded at different levels of resolution. However, using this approach, the user will only be provided with this visualisation having selected a cluster of interest.

[20] In later work Perry and Donath draw on Chernoff's faces - a multivariate scatter plot that makes use of the human features to map the attributes of specific data points (Chernoff, 1973) - to create Anthropomorphic visualisations, small visualisations that map user attributes to human-like features (Perry & Donath, 2004). While certainly novel, and arguably an interesting way to address multivariate community datasets, the approach produces a "caricature" of the user and the result, in some cases, may be irrelevant or even unacceptable. Also, there is no evaluation presented with the work, thus it is difficult to assess, or understand the general reaction to, this approach.

providing an engaging and rewarding experience for novice users. Based on an organisation's Computer-Mediated Communication (CMC) archive, the author's create both a thread and people visualisation that enables users to explore different aspects of community activity (see Figure 3). Again, retinal variables, such as colour, shape and size are used to represent users and threads. As with People Garden, a design implication that we can draw for this approach is based on the use of multivariate visual mappings to enable, what Tufte describes as, macro/micro readings (Tufte, 1990a). The user, when reading either People Garden or Communication Garden, is able to compare and contrast the activity of individual users but also develop an understanding of the community as a whole. However, as mentioned previously, comparative analysis can be difficult when addressing communities with a large population because, as suggested by the authors Zhu and Chen, a couple of hundred messages can appear as if there are one or two thousand in total. This is quite a limitation when considering the size of online communities and may contribute to the author's focus on organisational CMC archives. A second difficulty with this work is based on their approach to evaluation. The evaluation is task-based and the tasks are drawn from the information visualisation task-taxonomy originally proposed by Zhou and Feiner (Zhou & Feiner, 1998). Using a "de-featuring approach", the authors compare the visualisation to Netscape's Navigator program, a regular archive interface. The fact the author's approach out-performs the non-visual interface using tasks drawn from an information visualisation task taxonomy is hardly surprising and does not necessarily validate their approach. However, as with People Garden, the aim of communication garden is to provide a visual interface for novice users - essentially users that are not analysts or interested in conducting analytic tasks - thus the use of playful visual mappings can potentially broaden the audience and increase user engagement. Considering the use of playful mappings, coupled with the advantages and disadvantages of multivariate visual mappings, we propose the following two design considerations:

DC4. **Playful Visual Mappings**: Provide an engaging experience for casual users with playful visual mappings.

DC5. **Multivariate Visual Mappings**: Enable micro/macro readings with multivariate visual mappings.

3.3.2 Artistic Renderings

Some researchers have investigated the use of artistic renderings or, as defined by Viégas and Wattenberg, artistic data visualisation (Viégas & Wattenberg, 2007) to convey the social characteristics of different online community environments. Although these approaches may not provide the user with deep analytic insight, boyd et al. argue that artistic renderings are an effective way to convey the "sense" of a community's social space"[21] (Boyd, Ramage, & Donath, 2002). Lau and Vande Moere take a different view, arguing that artistic renderings have the potential to "entice user's interest over long periods of time" (Lau & Vande Moere, 2007). In either case, there is merit in considering how to address communities using approaches that attempt to balance artistic with analytic intent.

[21] A community's social space is essentially the community's environment. In this instance, the environment is Usenet, as discussed in chapter 2, one of the first discussion communities to reach internet scale. However, the use of the term "social space" describes the space as having a social dimension in the sense that it is dynamic and lived-in.

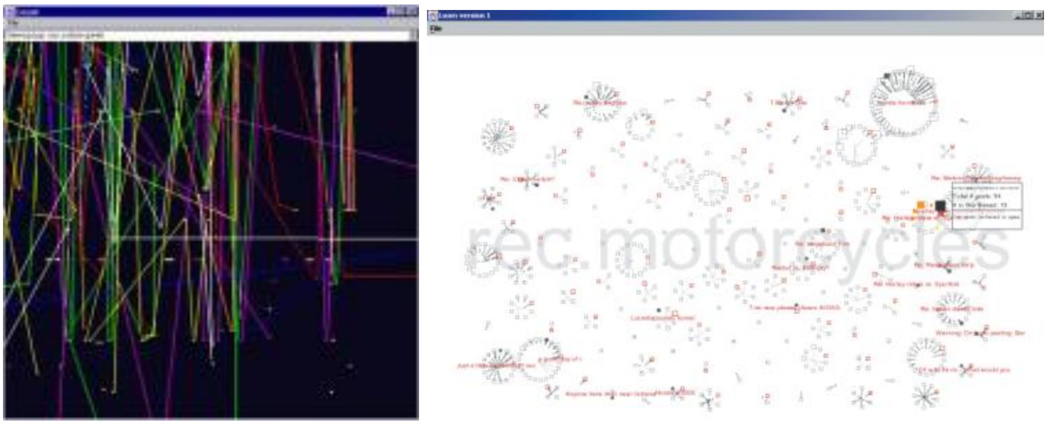


Figure 4: Artistic renderings from Loom (Donath, Karahalios, & Viégas, 1999)**. This first image depicts soc.culture.greek, an active and often argumentive newsgroup with rapid fire back-and-forth contests between different users. The second image represents the activity patterns of the Newsgroup rec.motorcycles.**

In Loom (Figure 4), for example, Donath et al. investigate how to create "legible" interfaces that represent a community's social space (Donath et al., 1999; Donath, 2002). The term "Loom" refers to how the system weaves "UseNet threads" into a rich digital fabric that reflects the "textures" and "patterns" of events, people and activity in Usenet[22]. The designers of Loom sought to support casual browsing by providing small "thumbnail" representations of different Usenet Newsgroups. This approach is similar to Tufte's small multiples, a popular way to analyse multivariate datasets (Tufte, 1990b) and discussed further in relation to online community datasets in Sect. 3.6.4. The resulting representation, however, is highly impressionistic, the coloured lines, darting both vertically and horizontally, represent the intense activity of a particular Usenet thread (see the first image in Figure 4). The second visualisation (see the second image in Figure 4) is more conventional in the sense that different retinal variables (shape and length in this instance, colour and length in the first example) are purposefully combined with spatial positioning to provide a more coherent representation of a particular Newsgroup. Both have function however. The first visualisation allows the user to quickly identify with the patterns of community activity, potentially enabling the user to identify argumentative threads in contrast to passive threads, while the second visualisation allows the user to compare and contrast user activity in a particular Newsgroup.

[22] Usenet was of the most prominent large scale community environments available at the time of Donath et al.'s work on Loom. Much of the early community visualisation work was designed to support the exploration of Usenet Newsgroups by analysts or casual users.

Figure 5: A visualisation generted using Loom2, this is a much denser display and designers attempt to evoke the tone of the social space. The tufts and scrapes that are visibile across the image resemble the scrapes that could be found in a skateboard park, their application illustrates that this is a social space that is in use.

Iterating on the original idea of Loom, the authors created Loom2. As with Loom, the authors draw on the notion of a "legible social landscape", where colour, shape and spatial arrangement are used to generate powerful yet identifiable mental images of different social environments (Boyd et al., 2002). The authors seek to create more "expressive" and compelling visual representations that have, unlike traditional approaches to information visualisation, subjective characteristics. To accomplish this, they propose a design framework that includes the use of *culturally sensitive visual metaphors*, such as red for stop and green for go, *a multi-scaled information visualisation*, so that users can zoom in-and-out of the visual landscape and, finally, *the use of kinetic typography*, or animated textual glyphs that visually communicate emotion. The result dynamically portrays the social characteristics - the patterns of activity coupled with the scuffs and scrapes of repeated social interaction – of different Usenet Newsgroups. While the image in Figure 5 may not provide much in terms of analytic insight, particularly for professional analysts or seasoned community users, the approach is effective at conveying a "sense" of the community's environment. The use of dark red coupled with texture and typography suggests a community in conflict, however, how this conflict is manifested is more difficult to ascertain. Thus, there is balance between analytic and artistic intent, which suggest the following design consideration:

DC6. **Artistic Renderings**: Convey the "sense" of an online community using artistic data renderings.

3.3.3 Things, Groups and Individuals

We can either represent communities at a system level - using statistical charts and aggregated data - or focus on visualising things, groups or individuals. We use the term "things" to describe objects that the community can create such as articles, blogs, comments and code, and how interacting around these objects can illustrate community dynamics. We consider this in terms of corpus visualisation as the aim is to visualise the activity of a corpus of content, which in turn, represents the dynamics of the community. We also address the visualisation of groups and individuals.

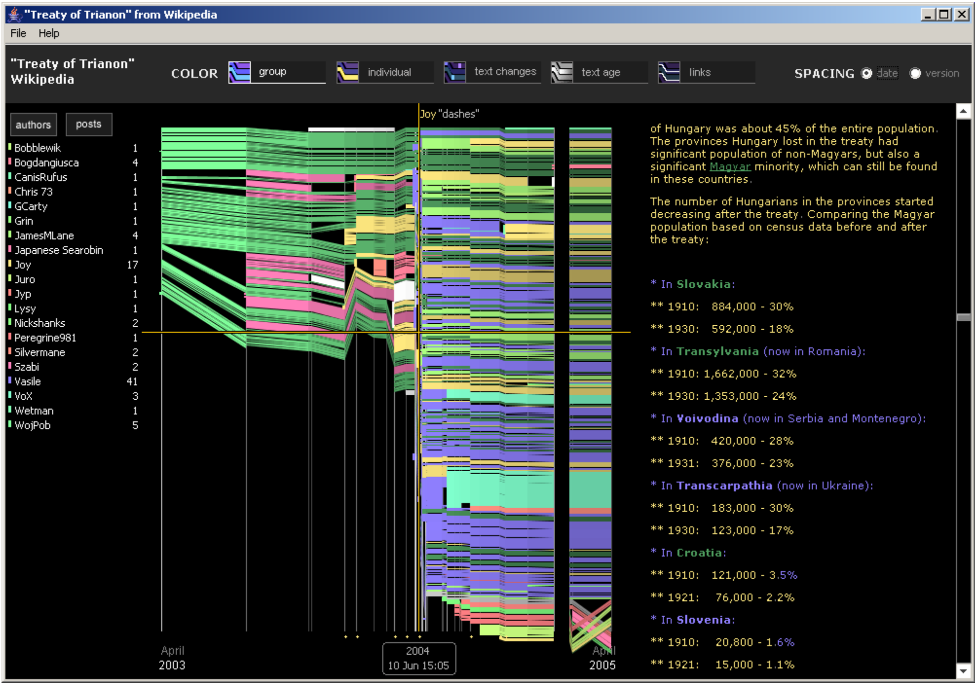


Figure 6: history flow diagram showing the edit activity on the Treaty of Trianon[23] article (Viégas, Wattenberg, & Dave, 2004; Viégas, Wattenberg, Kriss, et al., 2007). The most interesting aspects of the visualisation are the salient gaps, which illustrate mass revisions or mass deletions, and the zig-zag patterns, which illustrate edit wars, or continuous back-and-forth reverts, between users.

history flow, a visualisation designed by Viégas and Wattenberg, is an example of how to address the representation of "things" in online communities. With history flow, the authors explore revision activity in the English Wikipedia (Viégas, Wattenberg, et al., 2004; Viégas, Wattenberg, Kriss, et al., 2007) and were able to identify several important analytic insights, such as the edit war[24], and determine that the majority of vandalism is repaired within a two minute timeframe. Yet the composition of history flow, as illustrated in Figure 6, is quite complex. A revision to an article is represented by colour and then spaced by time. Sections of the text that remain the same between consecutive revisions are linked, while insertions and deletions lead to "visually salient gaps" in the resulting display. While the result is striking, particularly given the use of colour as a retinal variable, it can be difficult to read, especially for novice users. history flow, unlike several of the approaches discussed so far, was designed to assist in the inductive study of Wikipedia by the authors (as illustrated by the paper's findings). One design consideration that we can draw from history flow is the focus on representing activities that the community engages in as opposed to focusing on the community as a group (as illustrated with previous examples such as People Garden). Benkler describes the process of peer production in Wikipedia as involving both creation and maintenance phases - creation is the first phase in which a corpus of content is generated and maintenance is the second phase in which the corpus of content is refined (Benkler, 2006). This account holds for many other peer production communities, particularly content based communities but also open source communities hosted on SourceFordge and Github (considered in Sect. 3.5.3). By focusing on revision activity, history flow is adept at conveying community dynamics without emphasising the activity of a particular group or individual. We consider this a particularly effective social visualisation as it not only

[23] This is the English Wikipedia Treaty of Trianon History Flow.png, Wikimedia Commons, viewed 05th December 2012 and retrieved from http://commons.wikimedia.org/wiki/File:English_Wikipedia_Treaty_of_Trianon_History_Flow.png.

[24] Edit warring is the name given to situations in which adversarial users continuously revert each other's articles. Edit warring resulted in the introduction of the 3-Revert-Rule, see http://en.wikipedia.org/wiki/Wikipedia:Edit_warring#The_three-revert_rule, which stipulated that any user who commits three reverts in single 24 hours gets an immediate 24 hour ban. This is not an automated ban, however, but is placed at the discretion of Wikipedia administrators.

generates valuable analytic insight but also raises the curiosity of the reader yet remains aesthetically pleasing. history flow has since exhibited in MOMA[25], which is not only a testament to the approach's cultural reference but also the design's aesthetic appeal.

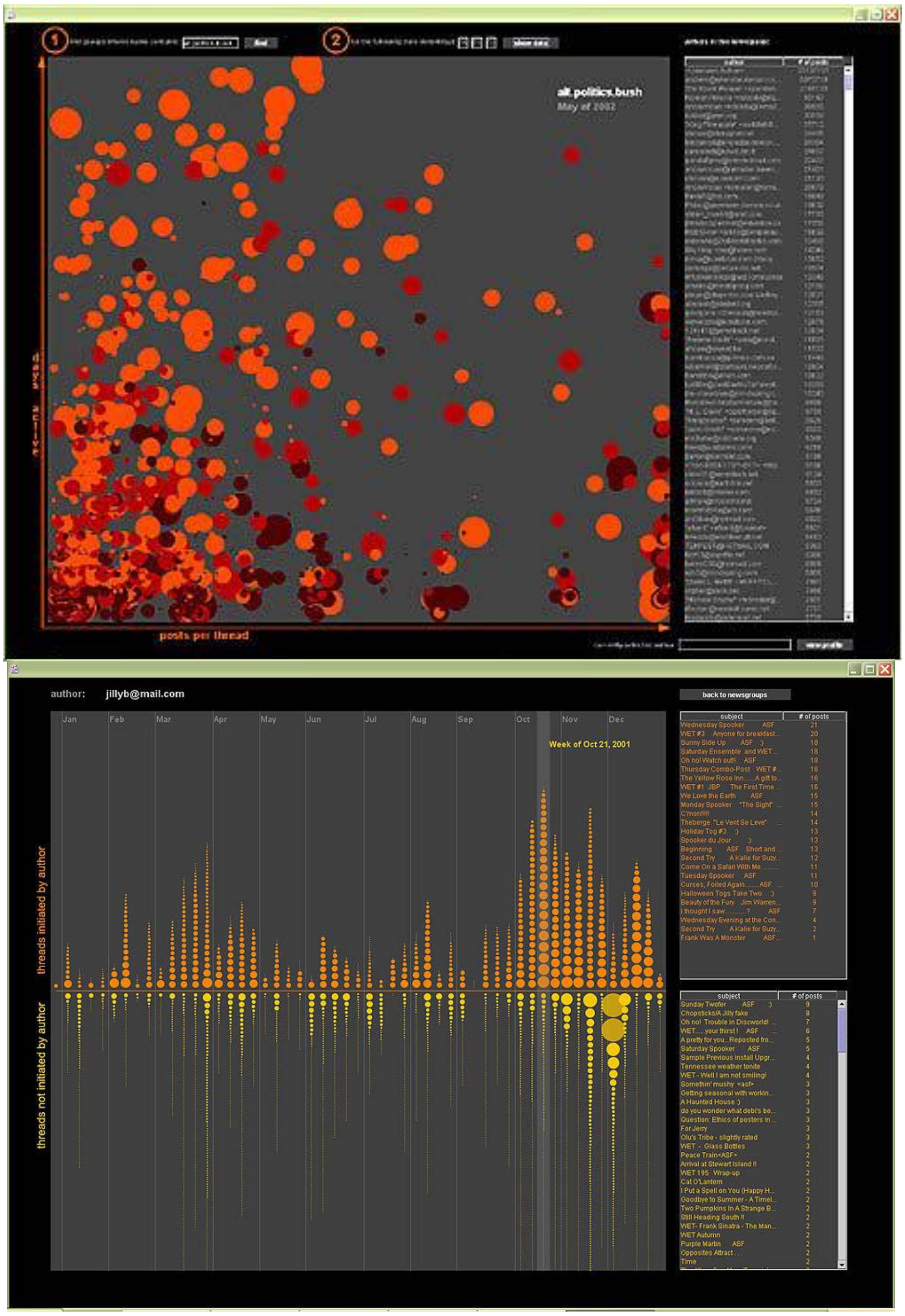


Figure 7: Two visualisations by Víegas and Smith, Newsgroup Crowds and AuthorLines (Viégas & Smith, 2004). Newsgroup crowds, the first image, is a multivariate scatter plot in which posting characteristcis are mapped to the spatial postiion, colour and size of each retinal variables and then plotted on a 2D scatter plot. Authorlines, inspired by LifeLines (Plaisant, Milash, Rose, Widoff, & Shneiderman, 1996) illustrates the post and reply behaviour of a single user over a period of time. AuthorLines enables the reader to quickly understand how a user contributes to a commuunity.

Also designed by Víegas are Newsgroup Crowds and AuthorLines, two visualisations that address Usenet at group and individual levels (Viégas & Smith, 2004). Unlike previous approaches discussed in Sect. 3.3, both Newsgroup

[25] See http://www.bewitched.com/historyflow.html.

Crowds and AuthorLines aim to accurately and objectively convey the activity patterns and contributory characteristics of users in Usenet. Newsgroup Crowds was the first visualisation to illustrate that patterns of recurrent activity are indicative of particular social roles in online communities (discussed in Sect. 3.2.2). This has since become a common research theme, explored by Smith et al. in numerous papers in which both Newsgroup Crowds and AuthorLines are used to illustrate the commonality of social roles across (mainly Q&A type) online communities (Turner, Smith, Fisher, & Welser, 2005; Welser, Gleave, Fisher, & Smith, 2008). AuthorLines, on the other hand, maps user activity over time. Although this approach makes it difficult to compare and contrast users, because the entire screen is used to visualise the activity of a single user, one can quickly surmise the sort of behaviour a user exhibits, whether, for example, the user posts more than they reply. Again, this is an effective way to consider user behaviour and to establish in what capacity a user contributes to a particular community. This analysis has shown that communities can be represented in numerous ways, each providing a different perspective on community dynamics and results in the following design consideration:

DC7. **Corpus Visualisation**: Convey community dynamics by visualising the activity of a corpus of community content.

3.3.4 Pixel-orientated Visual Mappings

A difficulty that we identified in Sect 3.3.1 was the ability of multivariate visual mappings – or visualisation representations that provide a one-to-one data mapping and enable macro/micro readings – to effectively address large online communities. As the community grows and the number of data points increase, the clarity of the visualisation representation is reduced. This is a considerable limitation given that online communities can have significantly large populations. One approach that can overcome this difficulty, yet retain the ability to address a large dataset, is based on the idea of a pixel-orientated visualisation. This approach, as outlined by both Keim and Shneiderman, involves mapping data points to individual pixels and then using colour to indicate attribute value (D. A. Keim, 2000; Shneiderman, 2008).

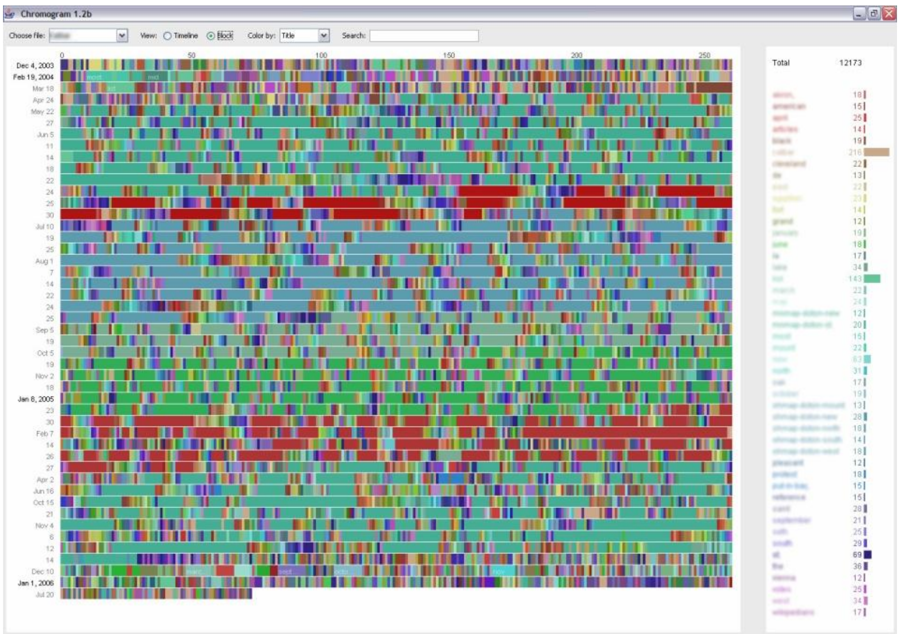


Figure 8: Chromogram Visualisation (Wattenberg et al., 2007). The authors use the first three letters of the revision string (the first three letter of a revision in Wikipedia characterises the type of revision) to represent the type of revision by hue, saturation and brightness respectively. The result is an information rich yet complex representation of edit activity in Wikipedia.

One approach to pixel-orientated visualisation, influenced by Keim's work on pixel-orientated visualisations (D. A. Keim, 2000) is Wattenberg et al.'s visualisation of editor activity in Wikipedia (Wattenberg et al., 2007). The aim of this work is to address an interesting question for social media researchers, how do editors allocate their time? Through the use of the visualisation, the authors found that editors prefer to focus on small yet reoccurring pieces of work, such as addressing a list of revisions ordered alphabetically, as opposed to switching from one large revision to another. However, the Chromogram, the name the authors give to the approach, is a specialist technique designed to enable the exploratory analysis of a large quantity of Wikipedia edit activity. The first three letters of each revision string is used to represent the hue, saturation and brightness of each Chromogram individually. The result, as illustrated in Figure 8, is a complex and dense visual display that can be a challenging to read and comprehend. Using the Chromogram technique beyond the present study warrants further investigation, particularly given our focus on users with little experience in the application of visualisation techniques. Moreover, the approach is not evaluated[26], nor are there many user studies in the literature on pixel-orientated displays. As illustrated in following papers (D. A. Keim, Schneidewind, & Sips, 2006; Riccardo Mazza, 2009; Stein, Wegener, & Schlieder, 2010), there remains scope to conduct more studies to ascertain the effectiveness of pixel-orientated visualisations when applied in different contexts.

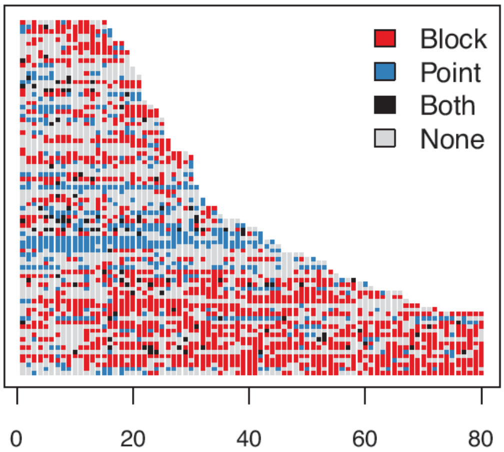


Figure 9: Chromogram Stack (Masli, Priedhorsky, & Terveen, 2011). This visualisation, which is suggestive of a heatmap, is a simplified version of the Chromogram visualisation developed by Watternberg et al. Different users are represented vertically and their activity (whether the revised a block on a map or a point on a map) is represented horizontally. Each coloured square is a type of revision. A reader of this visualisation can quickly establish how a particular user contributes revisions to the Wiki. From this perspective, it is much more legible and accessible than the approach presented in Figure 8.

The Chromogram technique has been applied elsewhere; however the composition has been simplified. Masli et al. use a variation of the Chromogram, which they call a "Chromogram stack", to explore the specialised revision activity of a geographic wiki for cyclists called Cyclopath (Masli et al., 2011). In this approach, each revision is assigned a colour - red for the revision of a block on a map and blue for the revision of a point on a map - and then user revisions are plotted horizontally. Using this visualisation, the authors find that authenticated users contribute more block revisions than anonymous users who tend to focus on point revisions. This is a much more accessible, and indeed legible, approach than the previous Chromogram visualisation (Figure 8) and the result is more familiar to readers, resembling the composition of a Heatmap. However, as with Wattenberg et al., the authors created the Chromogram stack to explore the revision activity of community users, thus there is no evaluation nor is there an intention to provide end-user support. In contemplating the advantages and disadvantages of pixel-orientated representations,

[26] This is not a critique of the study because this was not an objective of the study.

we suggest the following design consideration; however, we also acknowledge the need for the visualisation designer to remain cognizant of the user when using this technique:

DC8. **Pixel-orientated Visual Mappings**: Convey social patterns in large community datasets using pixel-orientated mappings.

3.3.5 Discussion

The challenge of aesthetics in Information Visualisation is to create representations that are engaging but do not reduce the legibility or function of the visualisation. A range of different approaches have been reviewed in this section, however some of the approaches are more appropriate to visual analysis than others. Artistic renderings, for example, lack convention, are open to interpretation - which is both culturally sensitive and context-dependent - and aim to generate reflective as opposed to analytic insight (Pousman et al., 2007). Accurate comparison of data points is often difficult, if not impossible, as is locating a particular point, a set of points or indeed identifying outliers. However, drawing on more artistic or playful visual mappings is not without merit. Research has found that some approaches, which make use of unusual aesthetic devices such as superfluous imagery - or "chart-junk" as described by Tufte - can in fact augment function. For example, Batemen et al. found that using the inverse of Tufte's design principles - increased chart junk and decreased data-to-ink ratio (Tufte, 1987) – did in fact improve rates of recall (Bateman et al., 2010). Daring and unusual designs can appear fun and engaging, attracting attention while improving rates of adoption, while scientific visualisation, which are designed to support analytic reasoning, can appear monotonous or in the words of Lev Manovich "anti-sublime" (Lau & Vande Moere, 2007). Further studies by Cawton and Vande Moere have shown that participants spend more time examining a visual representation when they consider the design aesthetically pleasing (Cawthon & Vande Moere, 2007).

Table 3: Evaluation of selected design considerations for aesthetic visualisation

	People Garden	Communication Garden	Loom & Loom 2	history flow	Authorlines NewsGroup Crowds	Chromogram
Playful Visual Mappings	**P**	**LE**	Ne	Ne	Ne	Ne
Multivariate Visual Mappings	**P**	**LE**	Ne	Ne	**US**	Ne
Artistic Visual Renderings	Ne	Ne	**P**	Ne	Ne	Ne
Corpus Visualisation	Ne	Ne	Ne	**AT**	Ne	Ne
Pixel-orientated Visual Mappings	Ne	Ne	Ne	Ne	Ne	**AT**

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not evaluated

A more general critique of this work is the lack of empirical evaluation. By cross referencing each design consideration with the approaches reviewed in this section, see Table 3, we can establish if and in what context each design consideration has been evaluated. It is clear there is a preference for demonstration over evaluation and the majority of proposed design considerations have not been addressed in user studies or lab experiments. Longitudinal case studies are considered a goal standard of information visualisation evaluation as the researchers can establish the utility of an approach in the context of use. However, in both instances of pixel-orientated visual mappings, the prototype was designed to assist the researcher's analysis of particular communities.

3.4 Mapping Community Networks

An approach that has grown in popularity, due partly to the proliferation of the internet and online social networks, is network visualisation. The application of network visualisation dates back to the early 1930s when Jacob L Moreno used simple hand-made drawings, called socio-grams, to chart the structural patterns of a small groups as illustrated in Figure 10 (Freeman, 2000; Trier, 2005).

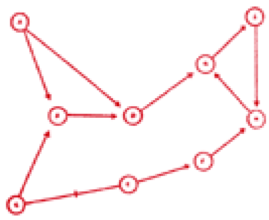


Figure 10: One of Moreno's first socio-grams charting who recognised whom amongst a collection of babies (Freeman, 2000). In this image, the network is directed, indicated by the arrows on each link.

The introduction of computers, coupled with an advance in algorithmic and visualisation techniques, accelerated the application of network analytics and enabled the investigation of increasingly large and diverse datasets. Online communities are particularly suited to this method of analysis because the community's structural features - the links and discourse patterns of the community - are persisted and often publicly available. Visualising these datasets has enabled researchers explore the relationship between repeated patterns of activity and identifiable social behaviours. For example, Welser et al. found that repeated structural signatures can be used to define the social characteristics of different social roles in community discussion environments, as illustrated in Figure 11 (Welser et al., 2008), while Kittur et al. developed a visual analytics system to assist in the identification of conflict in Wikipedia (Kittur, Suh, Pendleton, & Chi, 2007).

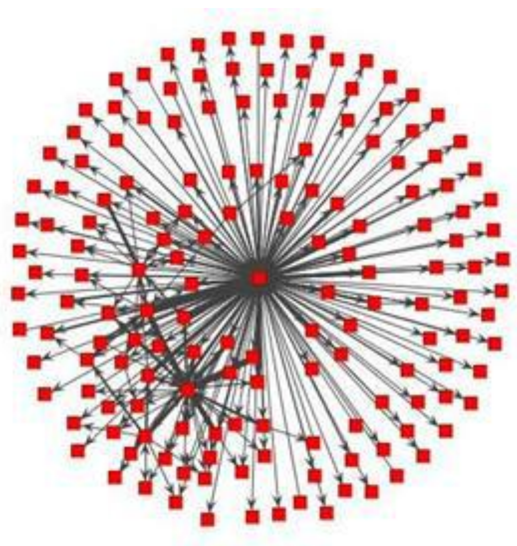


Figure 11: An ego centric node-link diagram showing a user's local network in Usenet (Welser et al., 2008). In this image the ego or central node is considered an "Answer Person" because they contribute to more threads than they initiate and Q&A is the community's common social focus.

In the majority of cases, the visual representation of networks make use of a technique called a node-link diagram in which nodes represent individuals and links represent connections between individuals (Fisher, 2005). In Figure 11, for example, connections from the ego or centre node are illustrated using a radial layout. This is a 2-degree directed ego-centric visualisation as connections between the ego's neighbours are also visible and the direction of correspondence is indicated. However, problems such as addressing scale-free networks have resulted in a range of approaches, techniques and systems. In this section, we review several of these. We focus particularly on how to design legible representations, on the relationship between network visualisation and the identification of social behaviour and also consider matrices as an alternative means of representation.

3.4.1 Spatial Layout

One of the principal challenges of network visualisation is how to layout the graph. Many visualisations that attempt to address networks without considering visual clarity can result in the derisively named hairball (Jeffrey Heer, Bostock, & Ogievetsky, 2010) - a tangle of nodes and links with little visual discernment as illustrated in Figure 12.

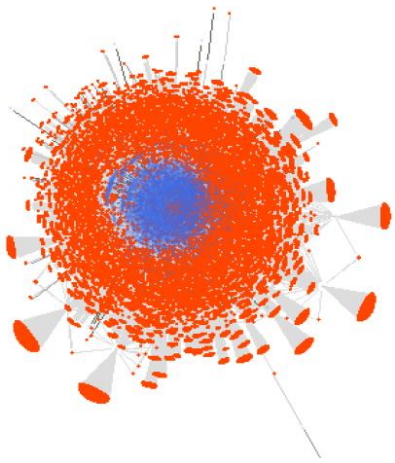


Figure 12: Bi-partite network visualisation of the role networker. The visualisation is typical of a "hairball" and needs further work to ensure a more useful and usable artefact (Modritscher & Taferner, 2012).

This is particularly a concern when addressing online communities or social media archives given the potential scale of these datasets. The effectiveness of a force directed layout algorithm - the popular network layout algorithm developed by Reingold & Fruchterma that dynamically creates clusters based on tightly connected nodes (Reingold & Fruchterman, 1991) - is dramatically reduced when addressing large networks. Furthermore, force directed layout algorithms use animation to organise the graph into a meaningful representation, which, when conducted on a large network, can have prohibitive execution times. One way in which a designer can circumvent this issue is to focus on the ego, as illustrated by Vizster discussed in Sect. 3.5, and enable the user to explore the network by traversing across nodes. The advantage of this approach is that large graphs can be explored interactively; however the user is unable get an overview of the network from a global perspective. An alternative approach, as proposed by Víegas and Donath (Viégas & Donath, 2004), suggests drawing on the work of Mark Lombardi, the American artist who creates precise hand-drawn graphs in which as links curve evenly around each node (Duncan, Eppstein, Goodrich, Kobourov, & Martin, 2010). Similar design principles have been proposed in the network visualisation community. For example, Ghoniem et al. outline several "aesthetic rules" that include minimising the number of crossed lines, minimising the ratio between the longest and shortest line and revealing symmetries (Ghoniem, Fekete, & Castagliola, 2004), while

Bonsignore et al. describe the state of "Netviz Nirvana " as having all nodes, links and clusters identifiable and countable[27] (Bonsignore et al., 2009). One design implication that we can draw from this work is to address network visualisation from the perspective of visual perception as opposed to mathematical computation. However, the scale of the dataset can still remain problematic. A recent approach, called SaNDVis by Perer et al., provides the user with a search and discover visual interface that helps users leverage social network information in the enterprise. SaNDVis is built on the SaND system, a social media mining platform, which generates a large graph of tags, documents and people based on the social media connection data of enterprise personnel. Having searched for a particular individual, the user is able to explore their relationships using the interface presented in Figure 13.

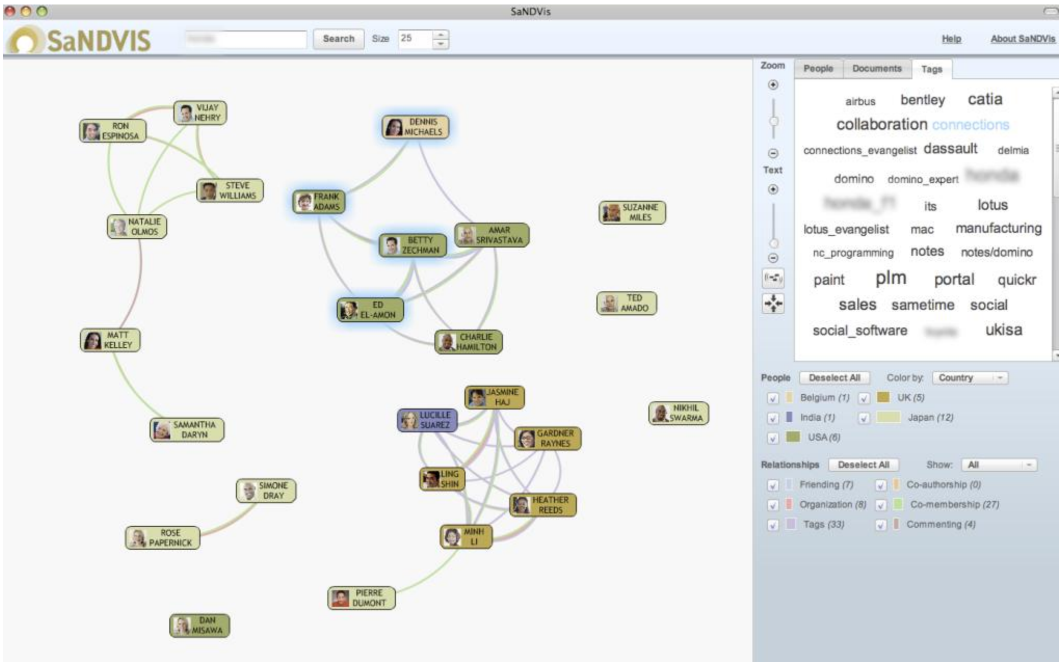


Figure 13: The SaNDVis interface (Perer, Guy, Uziel, Ronen, & Jacovi, 2011). The user is presented with a search bar that enables searching for particular enterprise personnel. Following a search, the user is presented with a tri-modal network visualisation, associated with tags and documents, to enable the exploration of enterprise relationships. The focus of the design on simplicity is of particular note, given the scale of the information being analysed and considering how the application is aimed at casual users.

What makes this system of particular interest is that, unlike SocialAction and NodeXL - two applications designed to support network analytics - there is no specialist knowledge required, in terms of network analytic or visualisation techniques. Rather, the aim is to allow casual users reflect on their own work relationships and the work relationships of their group (Perer et al., 2011; Perer, Guy, Uziel, Ronen, & Jacovi, 2012). This information can then be incorporated into how they coordinate their activities and develop their organisational practice. A second design consideration that we can draw from SaNDVis is the use of iconography. Icons have become a universally accepted way to represent users in social media or community-based systems. Both Vizster (Sect. 3.5.2) and NodeXL (Sect. 3.6.1), for example, make use of social media icons to label users in the visual interface. By focusing on the principles

[27] Other approaches have focused on attribute data as an organising principle. Wattenberg's multivariate graphs, for example, enable users to explore networks based on categorical data such as gender, age, hair colour etc. (Martin Wattenberg, 2006). Nodes are positioned and organised according to their attributes and then links are drawn between nodes. The result is much clearer with fewer overlapping links. Similarly, Shneiderman and Aris' approach, described as "network visualization by semantic substrates" uses attribute data to position nodes in a 2D space, similar to a scatter plot, and then draws links between those nodes (Ben Shneiderman & Aris, 2006). This approach has since been implemented in NodeXL. However, both multivariate graphs and Shneiderman usage of a semantic substrate are only appropriate for the analysis of datasets with attribute, and in the case Multivariate graphs. categorical data.

of visual perception, reducing the size of the potential network and using icons to label users, we draw the following two design considerations:

DC9. **Spatial Layout**: Represent node-link diagrams according to the principles of visual perception.

DC10. **Search and Discover:** Reduce the size of the dataset using a search and discover interface.

DC11. **Iconography**: Use social media icons to identify users.

3.4.2 Social Signatures

Smith et al. have investigated how visualisation can be used to identify "social signatures" – or the mapping of structural characteristics to social roles and processes (Butts, 2008) – in online communities. Initially, they focused on Usenet, however in later work, they broaden their analysis to include a range of different communities and shift their focus towards developing a more systematic methdology for the identification of social roles. Some of their first work used different visual representations to illustrate the structural characteristics of different Usenet Newsgroups (Turner et al., 2005).

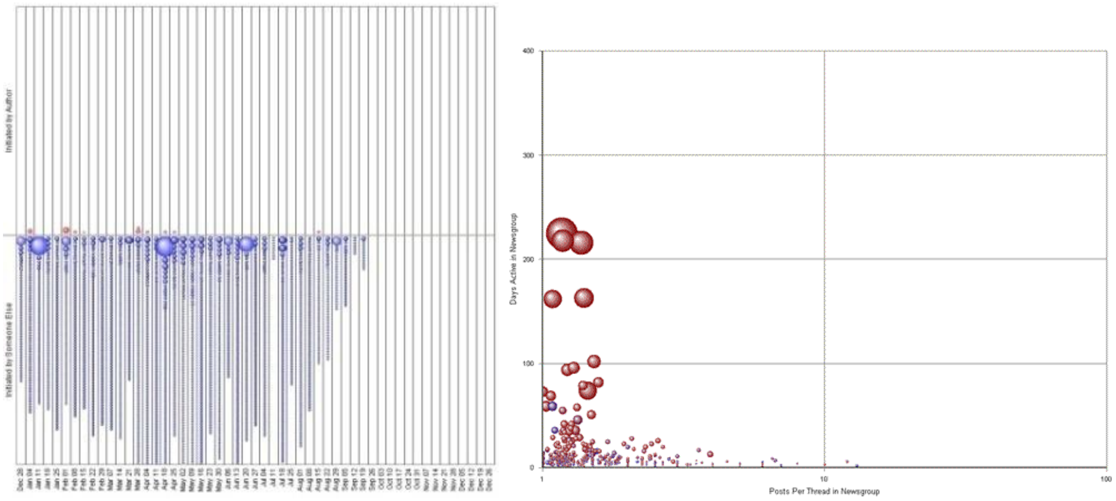


Figure 14: Two visualisations used by Turner et al. to explore the characteristics of individual and group behaviour in Usenet. The first visualisation, called AuthorLines, characterises users by their posting behaviour. The blue lines represent the number of answers (replies) a user has submitted while the red lines represent the number of questions (posts) the user has asked. This is the beginning of a structural signature for an "Answer Person". The second image, which is based on Newsgroup Crowds, plots users from the microsoft.public.windows.server.general newsgroup on a scatter plot according to their posting behaviour.

In Figure 14, for example, the authors use two visuisations, AuthorLines and Newsgroup Crowds discussed in Sect. 3.3.3, to investigate the behavioural characteristics of different users across different Usenet Newsgroups. Their analysis introduces the role of "Answer Person" whose "primary mode of interaction is the provision of helpful, informative responses to other group members' questions" (Welser et al., 2008). Technical support forums, and to a lesser extent discussion forums, tend to have a group of expert users that field questions from the rest of the user population. These users, as illustrated in Figure 14, have very distinctive contributory patterns because they rarely initiate but consistently contribute to discussion. One design consideration that we can draw from this is the ability for visualisation when mapped against repeated patterns of activity and interactive to illustrate social behaviour.

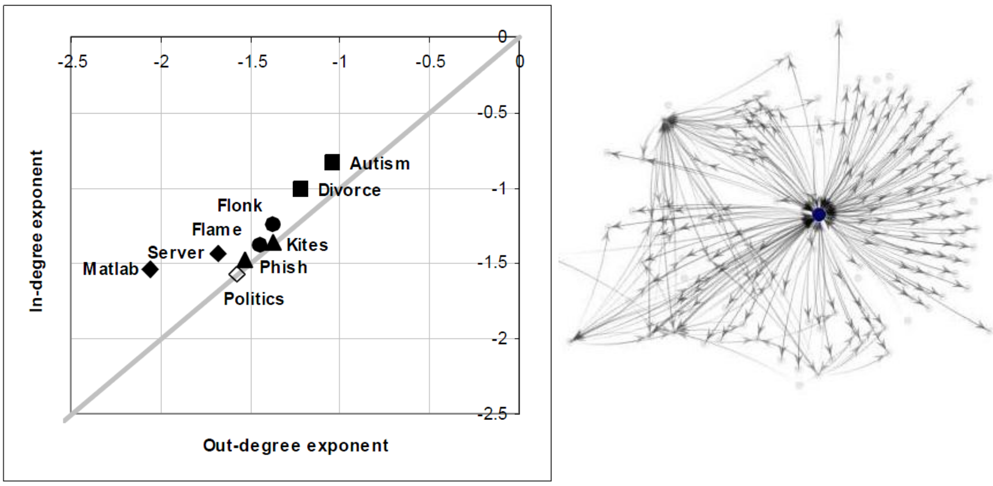


Figure 15: Scatter plot of power-law exponents indicates general behaviour in several newsgroups. Principally, the majority of users have similar degree distributions - their reply rate is approximately equal to their initiation rate. However, Matlab and Server have different distributions, their out-degree being less than their in-degree. This is consistent with a minority of users responding to a majority of questions. The second visualisation is a one-degree network representation of users in microsoft.public.windows.server.general. Notice how the majority of responses are provided by a minority (exactly two in this case) of users.

Of course, not all users have such distinctive structural signatures and in several instances user behaviour is closely related to the purpose of the community. This was shown in later studies where the authors address several additional newsgroups, including *alt.flame, alt.politics*, *microsoft.public.windows.server.general*, *comp.soft-sys.matlab* and *alt.support.divorce*, using a range of different visualisations including a scatter plot of exponent degree distributions and two degree ego-centric visualisations (illustrated in the first and second image in Figure 15)(Fisher, Smith, & Welser, 2006). As with their previous work, their findings indicate that certain Newsgroups, such as *microsoft.public.windows.server.general*, have very distinctive activity patterns, chiefly because Q&A is the common social focus, while behaviour in discussion communities is more nuanced and difficult to categorise.

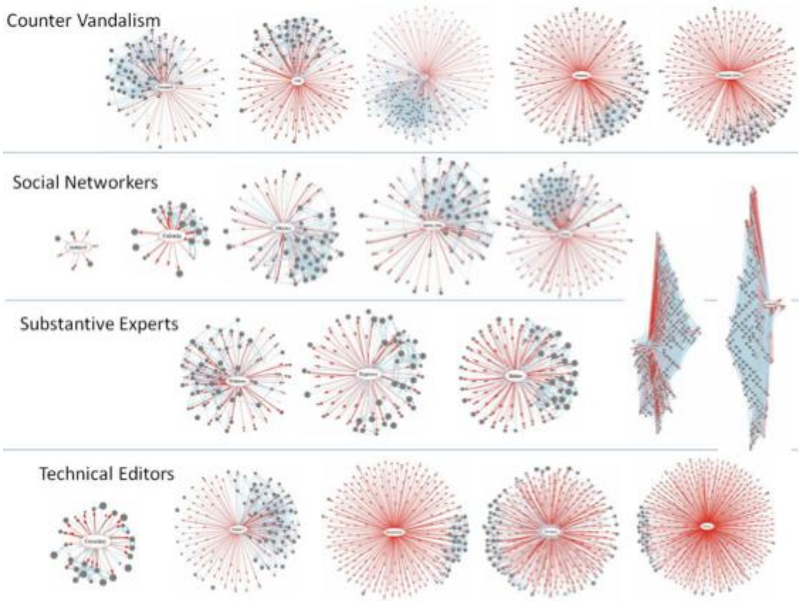


Figure 16: Structural signatures of different user roles in Wikipedia using ego-centric node-link diagrams (Gleave, Welser, Lento, & Smith, 2009). Although different social roles have structural similarities, it is clear from the visualisations that the structural characteristics of any single social role are much less obvious. This is a difficulty in social role identification, as often users will adopt different social roles at different times. Another challenge lies in the fact that different social roles adapt and change, making cross sectional analysis, as opposed to longitudinal analysis, much more challenging.

While this work illustrates the potential for visualisation to assist in the identification of social behaviour, Gleave et al. suggest that there is the potential for over-induction and over-deduction (Gleave et al., 2009). Over-induction results from the assumption that distinctive patterns of activity have a particular social significance while over-deduction results when we start with an abstract social category, such as altruism, and assume that particular patterns of activity fit within this category. In addressing this difficulty, the authors develop a systematic methodology to assist in the identification of social roles in online communities. Using Wikipedia, the authors demonstrate the process. First, drawing from the high-level categorisation proposed Kriplean et al.'s study of work in Wikipedia (Kriplean, Beschastnikh, & McDonald, 2008), the author's identify four key roles: Counter Vandal, Technical Editor, Social Networker and Substantive Expert (see Figure 16). Next, they take a directed sample of Wikipedia and iteratively traverse between content and structure to refine their understanding of each social role and validate the relationship between the role's structural characteristics and observable social behaviours. While the approach helps to systematise role identification in community discussion spaces, the results are much less substantive than the author's analysis of Usenet Q&A communities. Figure 16 illustrates ego-centric node-link diagrams of the four different social roles, Counter Vandal, Technical Editor, Social Networker and Substantive Expert. In this image none of the prescribed social roles have as distinct a structural signature as does the "Answer Person" illustrated in Figure 11. The social behaviour is much more nuanced and clearly different users adopt different behaviours at different times. Nevertheless, this work illustrates the ability of visualisation to not only illustrate social behaviour but also reveal the dynamics of online community environments.

DC12. **Social Signatures**: Indicate social behaviour by visually representing repeated patterns of interactivity.

DC13. **Reveal Socio-temporal Dynamics**: Use network visualisations to reveal socio-temporal dynamics.

3.4.3 Adjacency Matrices

Although node-link diagrams are the more popular way to represent networks, matrices provide designers with a useful alternative. With adjacency matrices[28], nodes are positioned on the horizontal and vertical axes and a connection between those nodes is represented at the intersection of those axes. Matrices can represent directed and weighted networks. Undirected networks can be composed by folding the matrix diagonally, which reduces space yet conveys the same amount of information to the reader. Directed networks, alternatively, require the entire matrix to illustrate connections between nodes. Weight can be conveyed via saturation, hue or intensity[29], so the number of connections between nodes is illustrated by the visual density of the pixels at the intersection of those nodes. The resulting representation conveys the same information value as a node-link diagram, as illustrated in Figure 17, but each approach has benefits and limitations.

[28] There are both adjacency and incidence matrices. Adjacency matrices, which we focus on in this section, are based on nodes that share a connection while incidence matrices are based connections that share a node.

[29] Using different colours is inappropriate as colour is more suited to representing categorical data as opposed to ordinal data. However, using saturation, hue or intensity can be less effective for users with impaired sight.

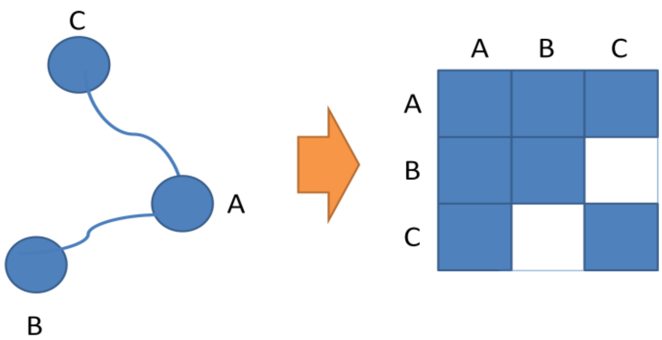


Figure 17: An illustration of the relationship between node-link diagrams and adjacency matrices. Both can encapsulate a variety of network configuration, provide the reader with the same information value however there are benefits and limitations to the use of each representation.

Researchers have investigated the use of matrices to represent a range of different social network data. Based on the outcome of several participative design workshops, Henry et al.'s developed MatrixExplorer (see Figure 18), a visual analytics environment designed to enable social scientists perform social network analysis.

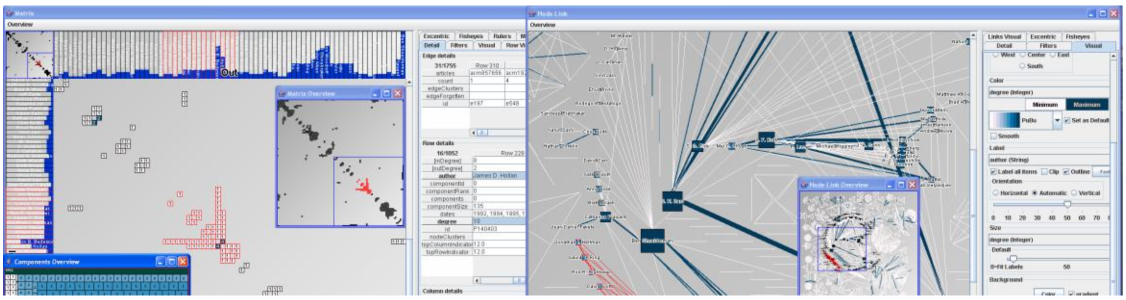


Figure 18: The dual representation - node link and matrices - of MatrixExplorer. Matrix diagrams are used to represent dense collections of nodes and links while node-link diagrams are used to illustrate less dense neighbourhoods, which in turn supports path-finding (Henry & Fekete, 2006).

MatrixExplorer is notable because, first, the authors include a group of users in the early stages of the design process. Furthermore, this group has varying levels of experience with network analytics and visualisation techniques Second, MatrixExplorer integrates matrices and node-link diagrams into a single visual interface (Henry & Fekete, 2006). One design consideration that can be drawn from this approach is the use of matrices to represent dense networks. In previous work Ghoniem et al. show that node-link diagrams are less readable when the network is dense[30] (due to factors such as occlusion), while matrices are much less useful when path-finding (the process of identifying paths between nodes)(Ghoniem et al., 2004). As illustrated in Figure 18, MatrixExplorer provides the user with a comprehensive analytic environment that enables the interactive analysis of large network datasets. NodeTrix, an extension of this work, increases the level of interactivity, enabling the user to manipulate large graphs, identify clusters of interests and alternate between representations in a seamless way (Henry, Fekete, & McGuffin, 2007). Although removing the need for programming experience, both MatrixExplorer and NodeTrix remain comprehensive analytic environments more suited to professional researchers than novice or casual users.

[30] Community networks are generally dense, scale-free networks with high degree hubs (Zhongbao & Changshui, 2003),

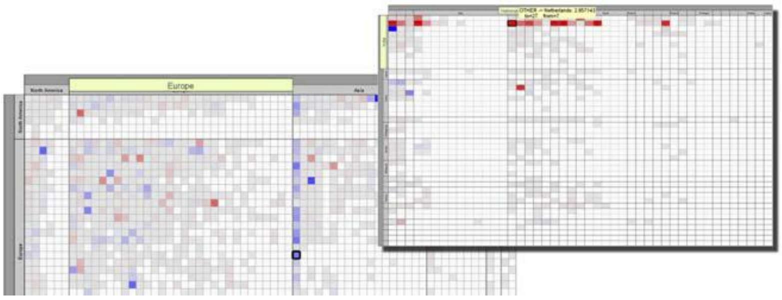


Figure 19: Honeycomb - a matrix diagram organised via country (Van Ham, Schulz, & Dimicco, 2009). This apoproach attempts to provide access to a large section of the entire IBM corporation. The aim is to provide concrete and recognisable organistaional hierarhices to drive analysis. This is successfull in that the approach is comprehnsible by the user while also providing a level of insight into how the communication patterns of the company.

Another approach that utilises adjacency matrices is Honeycomb by Van Ham et al. (Van Ham et al., 2009). Honeycomb is an explorative visual analytics system designed to handle large organisational datasets. The advantage of honeycomb is that it attempts to provide access to the entire network, which is unlike other systems, such as SocialAction (Sect. 3.6.1), SaNDVis (3.4.1) or NodeXL (3.6.1), that present the network from multiple perspectives, provide a snapshot of the network or couple statistics with both different visualisations. The authors argue that this approach places a lower cognitive burden on the user who has only to maintain a single mental map of the underlying dataset[31]. While this argument suggests that Honeycomb can be used by novice users, similar to the application of SaNDVis in Sect. 3.4.1, the system is never empirically evaluated so the claim is difficult to substantiate. However, we can draw several design considerations from this work. First, the authors use higher-order organisational abstractions, such as country or unit hierarchies, to drive analysis. This approach is in contrast to MatrixExplorer, which focuses on the use automatic clustering and ordering techniques, such as block-modelling, to structure the representation. The result of automatic clustering is often less intuitive than relatable abstractions, such as organisational hierarchies, and can, as argued by Scott, serve to confuse as opposed to support analysis (Scott, 1991). A second design consideration that we can draw from Van Ham et al.'s approach to Honeycomb is the ability for matrices to illustrate the absence as well as the presence of connections. A final design consideration, as suggested by Van Ham et al., is the ability for matrices to represent change more elegantly than node-link diagrams. This is because each node has a fixed position on a 2D plane and there is no need to use animation to maintain consistency when representing a change between nodes. This can be used to see network diffusion - the spreading of connections through a network. However, as discussed above, matrices are less proficient when "path-finding" - tracing the path between nodes. Although, the ability to path-find using node-link diagrams deteriorates significantly as the network's density increases as illustrated in (Ghoniem et al., 2004). Matrices are also less familiar than node-link diagrams and the majority of systems that address network visualisation focus on the node-link diagram as the primary means of representation. Nevertheless, matrices provide an alternative means of representation that hold several advantages over node-link diagrams, which we summarise in the following design considerations:

- DC14. **Dense Networks:** Represent dense networks with matrix diagrams.
- DC15. **Asymmetric Networks**: Reveal communication flows with asymmetric connections.
- DC16. **Absent Connections:** Show the absence of activity with matrix diagrams.
- DC17. **Higher-order abstractions:** Drive analysis using higher-order abstractions.

[31] Cognitive burden can be discussed in terms of cost structure. Cost structure is considered a combination of the "cognitive resources and external resources" used to complete an information rich task (Russell, Furnas, Stefik, Card, & Pirolli, 1993). It is important to consider the cost structure of an task and application particularly when addressing novice or casual users who have little experience of analytic tools or methodologies.

3.4.4 Discussion

From our analysis in this section, we have drawn the design considerations outlined in Table 4. As shown, a variety of techniques have been used to evaluate each design consideration, however, as identified in the other sections, there remains considerable scope to investigate their application in specific implementations. Interestingly, while matrices provide genuine benefits over node-link diagrams, their application is rarely considered. Finally, much of the work undertaken by Smith et al. sought to better understand how to map social structures to social behaviours, thus, this work is rarely evaluated from the perspective of users trial or lab experiments.

Table 4: Evaluation of selected design consideration for mapping community networks

	Multivariate Graphs	SaNDVis	Smith et al.	MatrixExplorer NodeTrix	HoneyComb
Spatial Layout	**US**	**LD**	Ne	Ne	Ne
Search and Discover	Ne	**LD**	Ne	Ne	Ne
Iconography	Ne	**LD**	Ne	Ne	Ne
Social Signatures	Ne	Ne	**AT**	Ne	Ne
Reveal Socio-temporal Dynamics	Ne	**LD**	**AT**	Ne	**P**
Dense Networks	Ne	Ne	Ne	**PDLE**	**P**
Asymmetric Networks	Ne	Ne	Ne	**PDLE**	**P**
Absent Connections	**US**	Na	Ne	**PDLE**	**P**
Higher-order abstractions	Ne	Ne	Ne	Ne	**P**

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not evaluated

3.5 Interaction and Animation

Interaction plays an increasingly important role in Information Visualisation. Soo et al. point to interaction as a way to address large datasets, overcome the limits of static representations and enhance a user's mental map by amplifying cognition (Ji Soo Yi et al., 2007). Drawing on well established approaches, such as Shneiderman's information seeking mantra (Shneiderman, 1996), the authors propose a categorisation of interaction based on user intent. Categories include select, explore, reconfigure, encode, elaborate, filter and connect. However, this categorisation can be broadened if we include coordinated visualisation and coordinated interaction mechanisms such as brushing and linking, drilldown and synchronised scrolling (North & Shneiderman, 2000). Many of these categories are found in the application of visual analytics[32] and in commercial analytic environments such as Tableau[33] and Chart.io[34]. Our interest in interaction, however, lies in providing ways to explore the socio-temporal patterns in online communities coupled with a focus on providing novice users with an engaging experience. Thus, we focus specificlly on how interaction can be used with animation. Heer argues that aimation is a double-edged sword; while facilitating constancy for changing objects and providing an emotionally engaging experience, there is the potential that animated behaviours to suggest casuality or intentionality (Jeffrey Heer, 2008). However, there are particualr benefits, first animation can be effective at showing the ebb and flow of communty life, which can improve perception and potentially increase a user's understaning of a community's socio-temporal patterns. Second, animation can be used to maintain orientation when transitioning between different states of a visual interface (Jeffrey Heer & Robertson, 2007). This can include exploration within a particular dataset of when transitioning between different visual representations.

3.5.1 Interactive Animation

The ebb and flow of online community life - the messages, the connections, the likes and periods of idle activity - make up rich social patterns that are obscured by text only interfaces. Visual Who, Judith Donath, was one of the first attempts to make these social patterns visible (Donath, 1995).

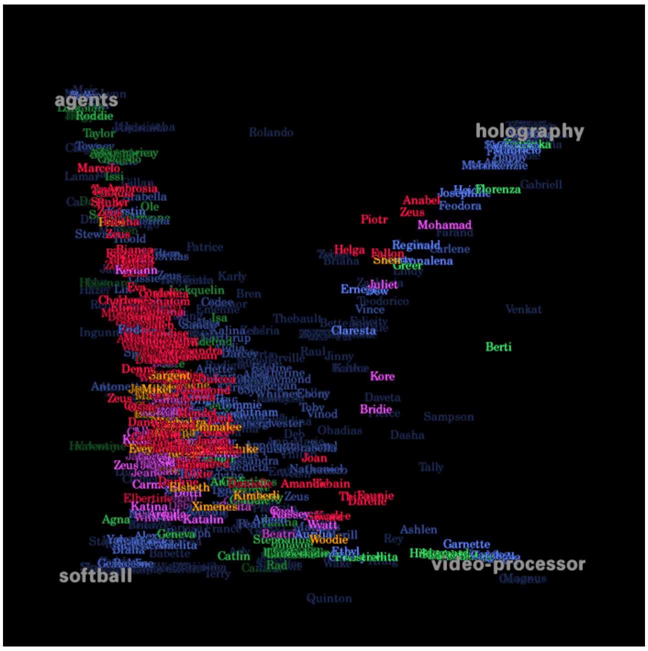


Figure 20: Visual Who: A visual "description" of a mailing list at M.I.T. The user can anchor different points and the visualisation, using animation and a spring layout algorithm, organises the community members into overlapping associated groups based on statistical analyses (described by the author's as the user's affinity) (Donath, 1995).

Visual Who, illustrated in Figure 20, is a visual description of a M.I.T Media Lab mailing list. The approach is based around the notion of "affinity". Users with similar characteristics, statistically speaking, are grouped together, while users that are unalike and have less "affinity" are placed at opposite ends of the visualisation. The visualisation is

---

[32] Interaction is one of the core research points outlined in Illuminating the Path, a research agenda for visual analytics (Thomas & Cook, 2005).

[33] www.tableausoftware.com

[34] chartio.com

then animated and the simulation enables users to see groups emerge and disperse. The animation, and the effect of group coherence, is further enhanced by the application of dynamic queries, which enables the user to set different anchors and thus reset the animation. Motion is particularly effective at conveying a sense of the ebb and flow of social groups, their interests and daily patterns of activity. However, visual analysis is more difficult. A user is able to explore the community and develop an appreciation of the community using this technique but not interrogate the data. In Figure 20, for example, it is difficult to identify the number of individuals with an interest in softball or video processing; however, we can quickly establish that softball is the more popular of the two. Not only is occlusion[35] an issue, but potentially, the animation reduces the accuracy of observation[36]. It is difficult to appreciate how the interactive animation was received by users, however, as the work fails to include an evaluation.

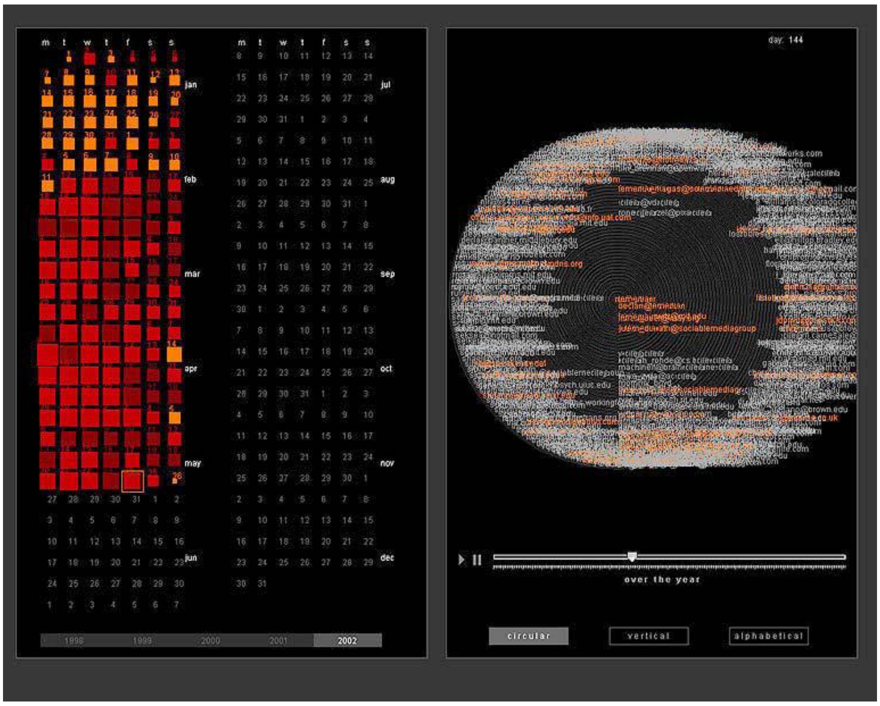


Figure 21: PostHistory - An email visualisation that enable the user to interactively explore and reflect upon their correspondence history (Viégas, Boyd, et al., 2004). The approach makes use of time as an organising principle and interactive animation as a way to convey social dynamics across time.

A second approach, which again makes use of interactive animation to illustrate socio-temporal patterns, is PostHistory[37] (Figure 21), an email visualisation developed by Viegas et al. (Viégas, Boyd, et al., 2004). The visual mapping is similar to Donath's approach to social visualisation (as illustrated with Visual Who) and, as can be seen in Figure 21, favours a more artistic rendering than accurate representation. PostHistory is based on the dyadic relationship embodied in typical email exchanges – the sender and the recipient. A large emphasis is placed on time, with calendar months used to break down the user's correspondence, and then a circular motif is used to illustrate how intensively a particular individual corresponds with the user. At the centre of PostHistory is the use of interactive animation to illustrate the rhythms – the ebb and flow – of daily email exchange. Unlike Visual Who, in

[35] Occlusion is when data points are obscured or hidden by other data points.

[36] The impact of animation on perception and the generation of insight in visualisation remains inconclusive (Archambault et al., 2010; Bederson, 1999; Tversky et al., 2002).

[37] Post History was developed with a second visualisation called Social Network Fragments but this second visualisation is not included in the analysis as it is beyond the scope of discussion.

PostHistory, the authors use an animated timeline. The passage of time is linked with the contacts panel, the second panel in Figure 21, and each contact sharpens in and out of focus based on the level of their email exchange with the user (or ego) at that particular point in time. Play and Pause buttons are provided, enabling the user to control the animation, an approach that has since been adopted by popular visualisations such as Gapminder[38]. Social relationships and group configurations naturally have a temporal dimension and animation is an evocative way to convey socio-temporal dynamics through the application of visual interfaces. This assertion is borne out in user trials in which the authors found the approach drew participants into periods of reflection and storytelling and that at several points during the trials participants indicated a desire to share the visualisation with their friends. There are some drawbacks to PostHistory however, most notably, the use of manual pre-processing to determine the types of interactions and the roles that different correspondents play in their lives. The need for a manual input restricts the application of the approach to archives of certain types and certain sizes, social media archives would become particularly problematic for example.

Figure 22: Bloom is a dynamic visualisation of activity in OSS projects (Kerr & Cheng, 2006). Based on a donut graph, the inner and outer rings represent code and comment contributions, illustrating who contributed to a project and in what capacity.

A third approach, called Bloom by Keer and Cheng, visualises contributory patterns to OSS (Open Source Software) projects hosted on SourceForge (Kerr & Cheng, 2006). The aim of the approach is to represent a project's activity so that new users and project managers can develop a "sense" of the project and identify key contributors. Bloom emphasises project communication in combination with repository activity. Thus, the authors visualise forum activity and other communications between project participants, along with the commit logs of the public software repository on SourceForge. They classify users into one-of-four of user roles according to a cursory analysis of their contribution patterns, the roles are "Coder", "Coder-Commenter", "Commenter-Coder" or "Commenter" and represent user activity as a donut chart (Figure 22). Again, interactive animation is used to convey activity. Similar to Post History, bursts of user activity emerge and contract in relation to change on a time-slider. These bursts of activity are represented by coloured circles extending outwards from the centre of the inner circle (see the inset in Figure 22). While the user can make out who is active during this time span, how active and in what capacity is more challenging. As with Visual Who, the use of animation can provide engaging experience but may be less effective for visual analysis. Despite numerous studies, including the following, (Bederson, 1999; Tversky, Bauer, Betrancourt, Europe, & St-martin, 2002), the degree to which animation can facilitate comprehension or "amplify cognition" remains an open question. Finally, despite its novelty, Bloom was never evaluated, making it difficult to empirically assess the utility of the approach. While it is arguable whether animation can result in better visual clarity, certainly the approaches

[38] www.gapminder.org

discussed in this section provide an evocative way to reflect socio-temporal patterns in groups and communities. Drawing from this analysis, we propose the following design consideration:

DC18. **Interactive Animation:** Use interactive animation to grab attention, increase engagement and show the evolution of socio-temporal patterns.

3.5.2 Interactive Exploration

A second approach to interaction uses animated transitions to maintain orientation during the exploration of large community datasets.

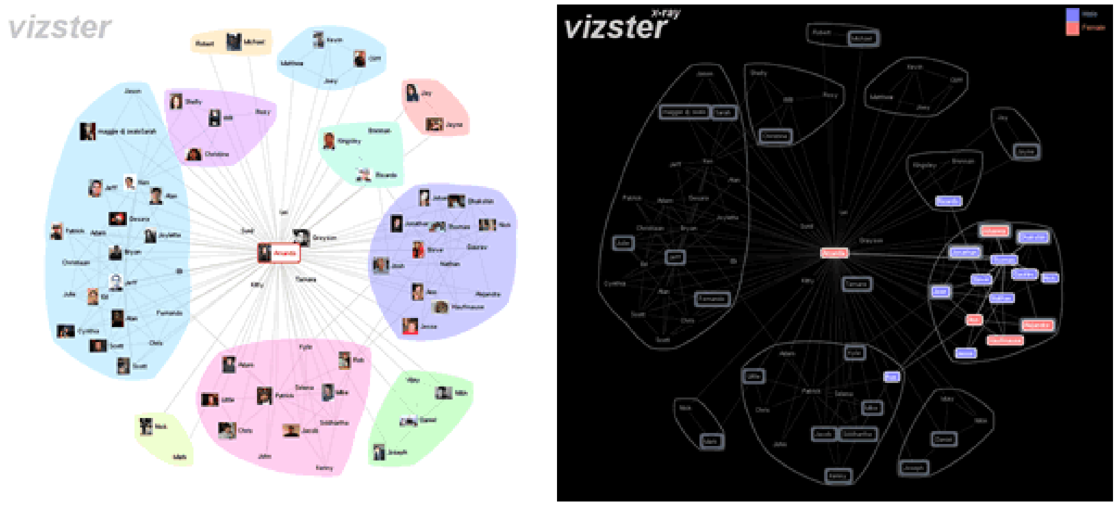


Figure 23: Two examples of visualisations produced by Vizster (Jeffery Heer & Boyd, 2005). The image on the left illustrates an ego-network, with the primary node in the centre of the image and related nodes scattered around the centre node. Related nodes are organised into clusters, or translucent blobs, according to Newman's link analysis algorithm. In the image on the right, a user is exploring the attributes (such as age, gender etc) of the different individuals using Vizster in x-ray mode.

Although there are a wide range of applications available that support network analysis, a large majority is designed to support expert users or professional network analysts. Vizster, in contrast, was designed to support "playful discovery" by the users of the social network Friendster (Jeffery Heer & Boyd, 2005). The approach seeks to remove many of the barriers associated with more complex analytic systems, such as JUNG[39] and GUESS[40], and in the process provide a fun and engaging way for casual users to explore their personal social networks. The approach is based on the application of ego-centric visualisation coupled with interactive exploration to maintain orientation. With ego-centric visualisation, the user is only privy to the community immediately surrounding a specific individual, or the ego, which restricts the ability of the user to develop a conceptualisation of the entire community. However, the authors circumvent this problem by implementing a form of interactive exploration in which the user, by clicking across nodes, can traverse their network. The layout of the network is computed in real-time and force simulation is left run to provide the user with a more playful and engaging experience. The tool was evaluated informally in the lab and, quite unusually, at a party in San Francisco. However, this is opportunistic evaluation strategy because the party was organised by users of Friendster and thus provided the author's with a relevant sample. The authors found that, as with PostHistory, participants engaged in reflection and story-telling. Onlookers also took part which illustrates, again, the potential for visualisation to promote social interaction. Unfortunately, further evaluation was not

[39] jung.sourceforge.net

[40] graphexploration.cond.org

conducted, nor was the tool deployed, so it is difficult to understand how the how interested users are in this sort of social visualisation or successful they considered the implementation. A second approach, which has been investigated by Heer and Roberston, is the use of animated transitions to maintain orientation when moving between different visual representations, for example, between a scatter plot and a bar chart. In lab experiments the authors found that that animation resulted in better graphical perception with significantly lower error rates (Jeffrey Heer & Robertson, 2007). The focus on providing an engaging experience, coupled with emphasis on exploration supported by animation, provides the following two design considerations:

DC19. **Interactive Exploration:** Use animation to maintain orientation when exploring large visual interfaces.

DC20. **Animated Transitions:** Use animated transitions to maintain orientation when moving between different visual representations.

3.5.3 Discussion

More recent studies investigating the benefits of interactive animation have provided contrary results as regards the usefulness of the technique. As discussed above, Heer and Roberston found that animation results in better graphical perception (Jeffrey Heer & Robertson, 2007), while Robertson et al., found that, in a different experiment, animation was confusing for users and small multiples proved best for analysis (Robertson, Fernandez, Fisher, Lee, & Stasko, 2008). At the same time, the authors recognise that animation provides an "exciting" way to present data in multiple dimensions and in many cases introducing new audience to Information Visualisation requires engaging visual interfaces. As with each design consideration discussed in this chapter, the use of interactive animation is a design trade-off and if unable to improve perception, the Google Maps and to a lesser extent Vizster, have shown that animated transitions can help maintain orientation during the exploration of large graphical interfaces. In Table 5 we identify how each of the above design considerations were implemented and evaluated.

Table 5: Evaluation of selected design considerations addressing interaction

	Visual Who	Bloom	Post History	Vizster	Heer & Roberston
Interactive Animation	**P**	**P**	**US**	Ne	Ne
Interactive Exploration	Ne	Ne	Ne	**US**	Ne
Animated Transitions	Ne	Ne	Ne	Ne	**LE**

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not Evaluated

As with many of the approaches discussed in Sect. 3.3, almost half the systems considered in this section were not evaluated. Furthermore, boyd and Heer's approach to the evaluation of Vizster, in which party goers casually browsed their social network using the tool, while unconventional, makes a degree of sense if we consider the purpose of the system. However, further lab studies would certainly be of benefit, or studies that focus on different aspects the tool, such how the approach to animation is perceived by users.

3.6 Integration and Configuration

Research has shown the potential for visualisation to amplify cognition is enhanced if designers can successfully integrate multiple visual representations into a single analytics system. Yet there is also the potential to overwhelm the user with complex visual interfaces and challenging interaction mechanisms. This is particularly the case when addressing users with little experience in the visual analysis of complex data. Successful approaches to social data analysis, such as Name Voyager and Wordle discussed in Sect. 3.7, are heavily orientated towards single visual widgets that provide users with a fun and interactive experience but can be readily shared using existing social media channels. However, the potential scale of community datasets, coupled with the ability for different visualisations to enable analysis from different perspective (as illustrated in Sect. 3.4.2) and at different levels of resolution, require the visualisation designer to consider integration and configuration.

3.6.1 Coordinated Visualisation

User-interface designers often coordinate several different visualisations into single interface to provide the user with a system that has more explorative and analytic capabilities. Coordinated visualisation, or sometimes described as Coordinated Multiple Views (Roberts, 2007), are based on a single interface within which is embedded a set of visualisations that are controlled by a set of coordination mechanisms, including brushing and linking, drilldown and synchronised scrolling (North & Shneiderman, 2000). The majority of systems that provide coordinated visualisations aim to support the explorative analysis of large complex datasets. For example, Social Action and NodeXL, two visual analytic environments that focus on the analysis of network data, both provide the user with a coordinated visualisation interface. One of the principal aims of SocialAction was to shift the use of visualisation from the end of the analyst's pipeline - in other words the presentation of findings - to the centre of the analyst's pipeline - in other words hypothesis formation and verification (Perer & Shneiderman, 2008). Inspired by Tukey's explorative data analysis (Tukey, 1980), the author's proposes a combination of different visual representations, presented to the user in conjunction with statistical information, to systematically assist in the analysis of (often large) network repositories. NodeXL could be considered an extension of this work but as opposed to focusing on the professional analyst the aim of NodeXL is to broaden the audience for network analytics by providing a more user friendly approach to the analysis of large network datasets[41] (Bonsignore et al., 2009; Smith, Hansen, & Gleave, 2009; Smith, Shneiderman, et al., 2009).

[41] Unlike other social network analysis packages, such as UCINet, Pajek and NetworkX, users require no programming expertise when using NodeXL, which is implemented as a plugin (or template) to Microsoft's popular Excel platform, and simple "sprockets" enable users to import data from Twitter, Facebook and other social media platforms. The system then provides an extensive array of statistical and visualisation techniques, and, in following Shneiderman's principles of Netviz Nirvana[41], users can generate a range of sophisticated network representations. However, careful assessment is required when generating a useful analytic pipeline. Thus, while removing the need to parse and organise data, the same methodological concerns arise when using NodeXL that exist with other less user-friendly packages. The user must understand how to apply the correct techniques in the correct order otherwise the resulting analysis is methodologically unsound. The use of recipes, common methodological procedures that can be reapplied on different network datasets, helps to mitigate these concerns, yet the problem remains when using comprehensive analytic systems.

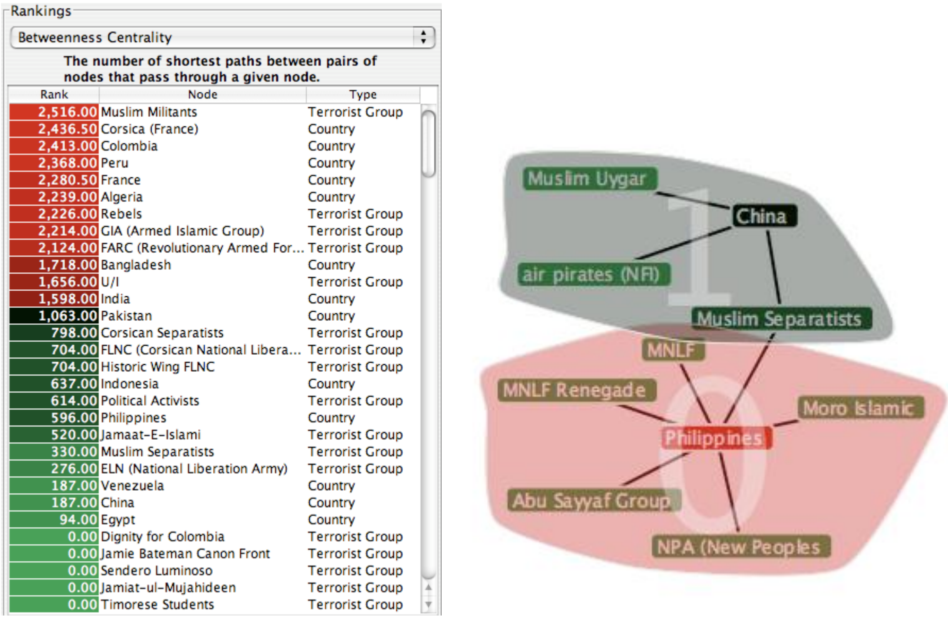


Figure 24: Two visualisations from SocialAction (Perer, 2006). SocialAction used ranked and coordinated views to enable the analysis of network data in a more systematic fashion. The user was provided with the ability to rank each node, based on a variety of different social network measures, and then to filter out different nodes using a double range slider. The result is presented as a coordinated view, a table coupled with a network visualisation, however, other visualisations, such as scatter plots, are also used to explore the network from multiple perspectives. The first is a tabulated view of nodes ranked according to a relative network measure. The second is a node-link diagram in which two separate communities (highlighted green and red) have been selected by the user for further examination. The aim of SocialAction is to cooridnate these displays, thus providing different ways to explore the data in a systematic fashion.

Both Social Action and NodeXL can be considered as comprehensive analytic environments however, that require users to have a considerable degree of expertise, not only in the application of visualisation techniques but also in terms of network analytics. A second approach, which can be considered by the visualisation designers, partly reduces the complexity of the interface yet enables the analysis of complex community datasets is illustrated with IChase, shown in Figure 25, an approach designed at Microsoft to support Wikipedia administrators (Riche, Lee, & Chevalier, 2010).

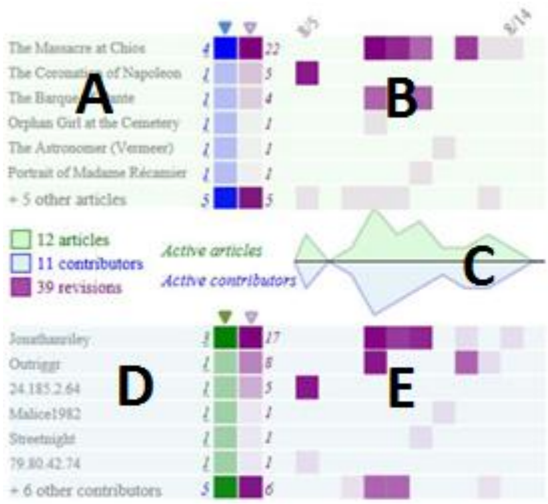


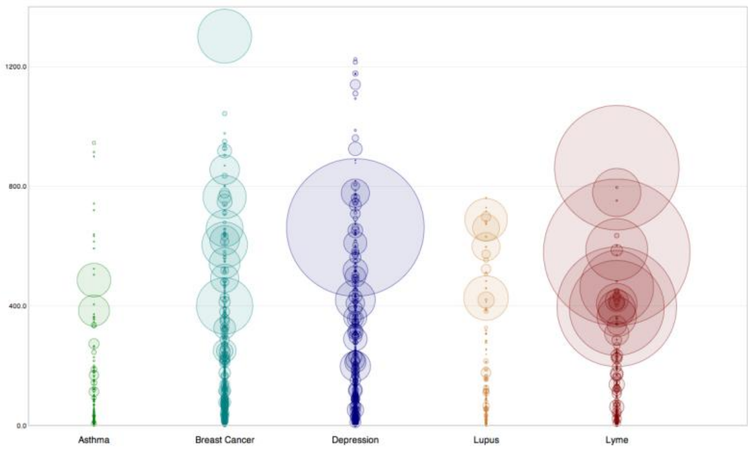
Figure 25: The IChase interface (Riche et al., 2010). A) Monitored articles B) The total contributions and revisions for each article C) A stacked graph showing article and author evolution D) Article contributors E) Author contributions.

Wikipedia, although conducive to authoring, reading and revising content, provides very little support for monitoring articles or maintaining system awareness[42]. To address this, Richie et al., in a similar approach to the design of CodeSaw discussed in Sect. 3.7.1, conduct a domain analysis to identify tasks that administrators find difficult with the existing Wikipedia interface and then, using partcipative design, develop IChase, a visualisation dashboard that assists administrators monitor article activity, identify vandalism and find mentors for new Wikipedia contributors. The design of IChase aggregates article, contributor and revision activity into a compact visual display that can be interactively navigated by users (Wikipedia administrators). This is a multi-scale (providing access to activity at different levels or resolution) and multi-perspective (enabling the analysis of different activities) visualisation dashboard that provides users with a composite visual representation of community behaviour. Unlike Social Action and NodeXL, IChase is not a workbench environment that requires users to apply an analytic pipeline before visual representation but, at the same time, provides users with a comprehensive overview of activity.

DC21. **Coordinated Visualisation**: Use coordinated visualisations to analyse complex datasets from multiple perspectives and at multiple levels of resolution.

3.6.2 Linked Visualisation

Some existing research suggests that coordinated visualisation can be difficult for novice users[43]. Learnability was shown as an issue, and often users resort to the most familiar yet inefficient means of analysis (Krishnamoorthy & North, 2005). As an alternative, linking different visual representations into a chain of related interfaces can reduce the complexity of system design yet enable the analysis of complex datasets from multiple perspectives.



[42] See http://en.wikipedia.org/wiki/Wikipedia:Toolserver for a list of Wikipedia tools. At present, the majority of tools are designed for more general purposes such as the identification of sock puppets or a list of administrators making difficult blocks.

[43] Considered in that particular study as users familiar with visual representation but less familiar with the process of visual analysis

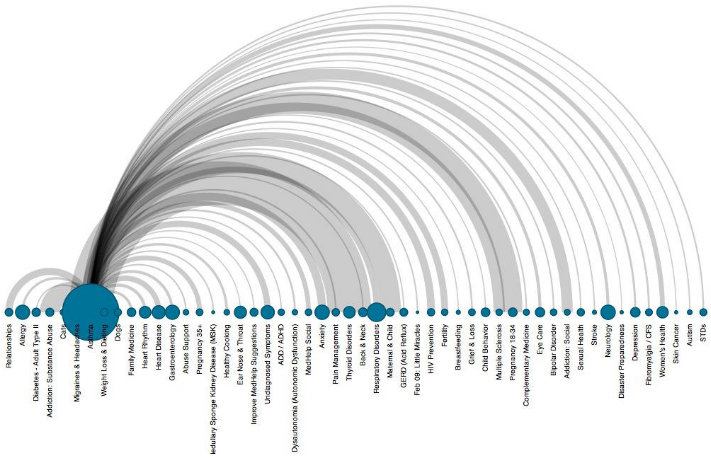


Figure 26: Two visualisations generated by MacLean and Hangal for their analysis of cancer support communities (Maclean & Hangal, 2009). The first, described as a "chromatograph", shows community role composition. The second is a co-occurrence visualisation, which shows different communities that users are visiting. Interestingly, this second representation can illustrate potential associations between different medical conditions.

Many of the papers reviewed in this chapter, particularly in Sect. 3.3.2 and Sect. 3.4.2, use several visualisations to convey different elements of a particular social phenomenon. These visualisations are not coordinated but are associated or linked together. Linking visualisation in this way reduces the complexity of the user interface yet enables the analysis of community at different resolutions and from different perspectives. A more recent example of linking visualisations is Maclean and Hangal's work on health care communities. Using several visualisation techniques, two of which are illustrated in Figure 26, the authors explored the social dynamics of cancer support communities (Maclean & Hangal, 2009) and found that patterns of visitation to different communities showed potential associations between different medical conditions. Linked visualisations can reduce the speed of analysis, as the user must traverse between different visualisations yet maintain orientation and a revised mental map. However, as discussed in Sect. 3.5, interactive animation can be used to maintain orientation when transitioning between different visual representations.

DC22. **Linked Visualisation:** Enable analysis yet reduce complex interface design with linked visualisations.

3.6.3 Embedded Visualisation

Another consideration for visualisation designers is whether to embed the visualisation in the community's existing social interface or as a visual widget within a corpus of community content. Some research, such as Erickson and Kellogg's work on social translucency (Erickson & Kellogg, 2000) and Suh et al.'s work on social transparency (Suh, Chi, Kittur, & Pendleton, 2008), has shown that embedding visualisation in social and community interfaces can improve collaboration and support end-user analysis. For example, Wikidashboard, illustrated in Figure 20, was shown under experimental conditions to improve the perceived credibility of Wikipedia articles (Pirolli, Wollny, & Suh, 2009).

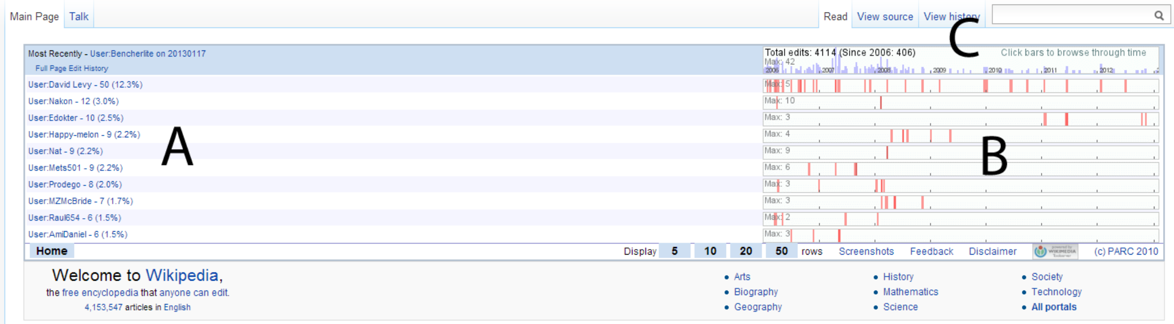


Figure 27: Wikidashboard as embedded in a Wikipedia page. A) List of Article contributors, including the percentage of contributions they have made to the article B) Weekly editor activity, the more saturated the red the more edits an editor has performed C) Weekly edit trend of this page.

However, more recent work has focused on the analysis of multiple communities simultaneously yet emphasised the importance of embedding visualisation within community content. CommunityCompare, by Xu et al., uses inetractive parralel coordinates to compare the performance of corporate sponsored online communiteis (Xu et al., 2013). The ability to directly access the content of the community was the most valued feature in preliminary user trials. This approach was first introduced with Sack's Conversation Map, see Figure 28, and enables users of a visual interface to verify their observations with reference to the actual content (Sack, 2000b). A second design consideration that we can draw from Xu et al.'s work, which was also addressed by Maclean & Hangal's approach to the visual representation of cancer support communities in Sect. 3.6.2, is the ability to compare and contrast different communites using relatvely simple visualisation techniques. Approaches applied across communities can potentially provide more insight than an approach applied to a single community. This is because a single community focuses the analysis on an individual case, while multiple communities provide a context for comparison. This is not only evident in Xu et al.'s work, but also illustrated in the increasing number of papers that analyse several communities simultaneously, such as Furtado et al.'s recent analysis of Stack Exchange (Furtado & Andrade, 2013). Considering the relationship between content and visualisation, and focusing our analysis across communities, provides us with the following two design considerations:

DC23. **Embedded Visualisation:** Enable verification and validation by associating visualisation with the underlying content.

DC24. **Cross Community Analysis:** Provide a context for comparative analysis by representing different communities within a single visual interface.

3.6.4 Small Multiples

Tufte popularised the notion of small multiples (Tufte, 1990b) - several small visual representations presented in a grid - although the idea was first introduced by Bertin and described as "collections". While particularly effective at exploring multivariate datasets, small multiples have also been used by visualisation designers to show different aspects of online community environments, as illustrate with Loom in Sect. 3.3.2, and illustrate change within complex community networks (Archambault, Purchase, & Pinaud, 2010).

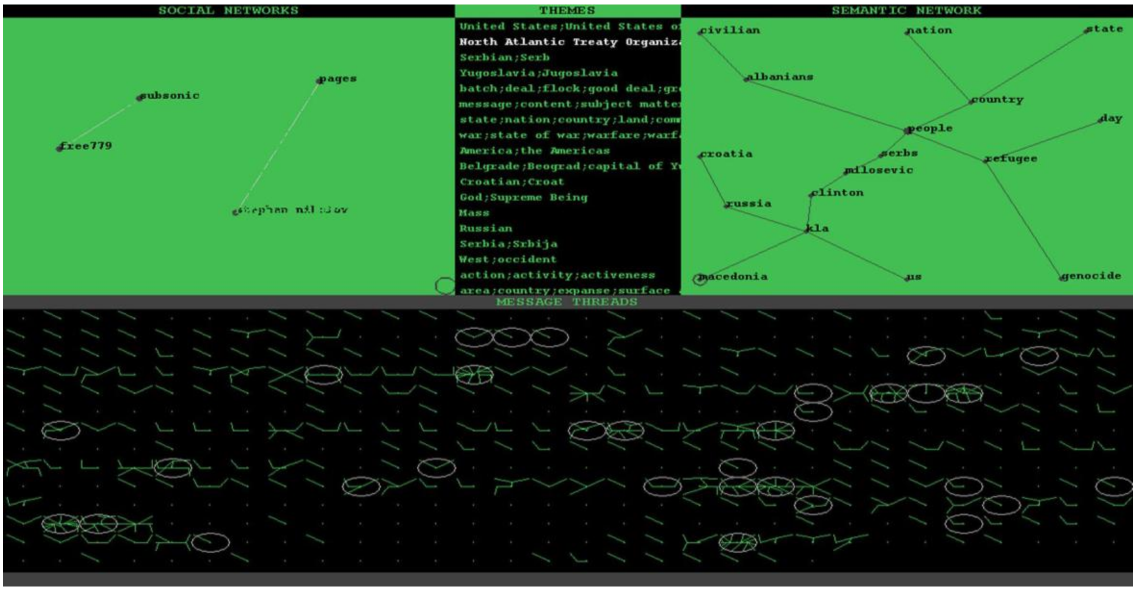


Figure 28: Conversation Map has both social and semantic visualisations (Sack, 2000b). On the right is the semantic network, linking the different concepts from a Usenet message thread. On the left is the corresponding social network, linking users who contributed to that particular message thread. In the centre are the discussion themes of the conversations, which are approximated from social and lexical analysis. The small networks aligned underneath are graphical representations of threads ordered chronologically.

Sack's Conversation Map, for example, was one of the first community visualisations that made use of the small multiple technique (Sack, 2000b). At the bottom of Figure 28, each thread is allotted a small rectangle and then organised chronologically. This enables the user to quickly compare and contrast threads by their structural characteristics, given that threads can have single or several branches. Donath's approach to Loom (Donath et al., 1999), as discussed in Sect. 3.3.2, along with Wattenberg and Millen's approach to Conversation Thumbnails (Wattenberg & Millen, 2003), also utilise this technique by presenting the user with small visual representations of discussion threads. Both of these approaches provide the user with an understanding of the community's conversational space, enabling the user to identify active from passive threads. Another benefit of small multiples that we can draw from this approach, and which has received some attention from researchers in graph visualisation, is the use of small multiples to represent the evolution of a community network. In a formal study, Archambault et al. found, for example, that small multiples are more effective in preserving a user's mental map, when analysing the evolution of a network, than the use of animation (Archambault et al., 2010).

DC25. **Small Multiples**: Enable comparative analysis of users, groups or communities, or demonstrate the evolution of an online community using small visualisations repeated in a grid.

3.6.5 Discussion

There are advantages and disadvantages for each of the approaches discussed in this section. Coordinated visualisation, for example, can enable the analysis of complex datasets but also potentially overwhelm novice users while linked visualisations require users to retain and revise their mental map. As with the majority of design considerations developed in this chapter, each approach suggests a trade-off and requires the visualisation designer to think about the intended user community and consider how the system will be used in practice. In terms of evaluation, as shown in Table 6, a range of approaches have been employed. In the context of the design of visualisations that address a particular user group, participative design followed by user studies conducted in context appear to be favoured. Certainly, recruiting larger sample sizes and finding systems for comparison is beyond the scope of many of these studies, while participative design coupled with user studies allows the system to be studied in context with the correct users, which potentially yields more qualitative insights.

Table 6: Evaluation of selected design considerations addressing integration

	SocialAction	NodeXL	IChase	MacLean and Hangal	CommunityCompare	Conversation Map	Conversation Thumbs
Coordinated Visualisation	**LD**	**LD**	**PDUS**	Ne	Ne	**P**	Ne
Linked Visualisation	Ne	Ne	Ne	**AT**	Ne	Ne	Ne
Embedded Visualisation	Ne	Ne	Ne	Ne	**US**	**P**	**P**
Cross Community Analysis	Ne	Ne	Ne	**AT**	**US**	Ne	Ne
Small Multiples	Na	**LD**	Ne	Ne	Ne	**P**	**P**

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not Evaluated

3.7 Collaborative Support

We include a category on collaboration because community design can be a collaborative endeavour and, as illustrated by the increasing interest in casual visualisation and social data analysis[44], web users increasingly share and discuss visual media. However, there is a notable lack of collaborative support in the community and social visualisation literature. Given the collaborative nature of online communities, coupled with the advances in web-based visualisation (most notably the JavaScript framework D3), it is surprising that more approaches do not incorporate some form of collaboration in their design. This may relate to the maturity of the field, as many of the approaches reviewed in this chapter never result in actual systems that can are evaluated in "ecologically valid scenarios". Our consideration of collaborative support is based on simplicity and ease of use, and in providing intuitive ways to encourage and support collaborative visual analysis.

3.7.1 Annotation

When collaborating around visual media, Heer and Agrawala argue that it is common to point to objects or regions to help emphasis something of interest (Jeffrey Heer & Agrawala, 2008). This level of reference enables grounding and grounding is required for individuals to share a common understanding of a particular problem or scenario. In asynchronous collaborative environments, annotations, such comments and tags, are typically used to assist in grounding. However, these annotations are external to the visual media and thus will lack a degree of accuracy. An alternative approach is to embed annotations directly on the visualisation. This can include embedded comments or

[44] Social data analysis is the term used to describe the use of interactive visualisations coupled with annotation tools to enable analysis by data analysis by broad user communities. This approach has been investigated by Heer, Viégas and Wattenberg and is discussed in (Jeffery Heer et al., 2009; Jeffrey Heer & Hellerstein, 2009; Wattenberg, 2005).

graphical annotations. Gilbert and Karahalios' CodeSaw, an approach drawn from the social visualisation literature, allows users to embed comments directly on the visual interface (a technique they describe as spatial messaging) (Gilbert & Karahalios, 2007, 2009). CodeSaw[45], as shown in Figure 29, is an interactive visualisation designed to enable small OSS development communities to reflect on community dynamics. Drawing on the design of Bloom (Sect. 3.5.1), the authors created a compact representation of code and comment contributions to OSS projects on SourceForge.

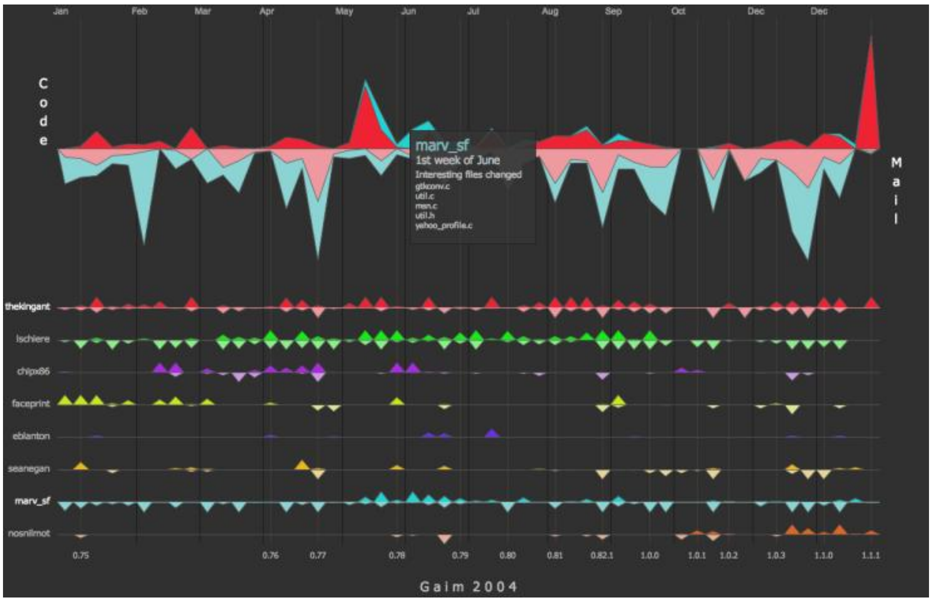


Figure 29: CodeSaw: The repository contributions are visualised above the x-axis and mail submissions are visualised below the x-axis communities (Gilbert & Karahalios, 2007, 2009). In the top panel, different user contributions are mapped against each other. In the bottom panel, small multiples are used to illustrate the contribution activity of eight different users.

The aim of spatial messaging is not to replace the accepted channels of communication but to enable the identification of points of interest in the visual representation. Users are able to associate a message with a particular state in the visualisation - a point in time for example and the annotation appear when user interactively traverses to that point. Alternatively, graphical annotation can enable users to identify, highlight and emphasise points of interest or, indeed, provide a way for users to engage in more playful analysis and commentary. From this perspective, Heer et al. developed Sense.US to evaluate their design approach to social data analysis (Jeffery Heer, Viégas, & Wattenberg, 2009).. Sense.US is a web-based visualisation environment that enabled groups of casual users at IBM to explore American census data The design of Sense.US includes several graphical annotation tools that act like a "acetate layer over the visualisation" and enable users to draw freeform or add geometric annotations directly on the visualisation state. Both approaches provide a more accurate means of reference but can also be used in a more playful capacity. Aside from annotation, a second design implication that we can draw from the approaches discussed in this section is their focus on simplicity. The designers of CodeSaw, for example, found that users preferred simple intuitive designs as opposed to approaches based on complex measurements. Participative design workshops followed by iterative design cycles, enabled the authors to reduce the design of CodeSaw into a necessary set of

[45] The composition of CodeSaw is also noteworthy. The approach, a variation of the stacked graph (Jeffrey Heer et al., 2010), is more legible than Bloom. Each user's code contributions (code committed to the repository) are placed above the x-axis and their mail contributions are placed below the x-axis. The user can quickly see three things: who are the principle contributors, in what capacity do they contribute (code or email messages) and at what periods during the project are they active.

design elements. We can draw the following two design considerations from both the design of CodeSaw and Sense.US:

DC26. **In-view Annotation**: Enable users to reference objects in the visualisation and encourage commentary with in-view annotation.

DC27. **Simplicity**: Support novice users with simple interface design.

3.7.2 Sharing

Some existing research has shown that visualisation has the ability to encourage sharing. Víegas and Smith, for example, found that when evaluating PostHistory, discussed in Sect. 3.5.1, participants indicated a desire to share the visualisation with their friends and colleagues. Similarly, Wattenberg found with Name Voyager - a simple yet engaging interactive visualisation downloaded almost a million times - that users were naturally inclined to share the visualisation with friends and reference specific states in the visual representation using comments (Wattenberg, 2005). Similar social behaviours where found with the application of Wordle (Viégas, Wattenberg, & Feinberg, 2009), while with Many Eyes, Danis et al. found that several users preferred to share visualisations from Many Eyes within their existing social space as opposed to engaging in discussion on the Many Eyes website (Danis, Viégas, Wattenberg, & Kriss, 2008). We can draw several design considerations from this work. First, each approach can be shared and referenced. In some cases, reference involves graphical annotation while, in others, comments enable users to identify points of interest and share the representation's state (as discussed in Sect. 3.7.1). Sharing supports grounding by enabling users to consider the same piece of visual media but also encourages further analysis by allowing users to share the visualisation with their friends or colleagues. Second, in each case, interactive animation is used to provide users with a fun and engaging experience (this is discussed in more detail in Sect. 3.5.1). Third, there is a preference for visualising personally relevant data. On this final point, Heer and Agrawala suggest that personally relevant data can increase usage rates due to "hedonistic incentives" (Jeffrey Heer & Agrawala, 2008) or, in other words, data that is of intrinsic interest to the individual or group. From this analysis, we point to the following two design considerations:

DC28. **Sharing**: Encourage collaborative analysis and enable grounding with URL or in application sharing.

DC29. **Personal Relevancy:** Use personally relevant data to drive analysis.

3.7.3 Commentary

Another consideration for designers is how to address discussion. The method that has proven most popular on the web, due most likely to its simplicity, is based on short textual communications linked chronologically or as replies in a threaded discussion. Yet, while generally accepted, this approach is partly flawed as the comments fail to accurately capture the pragmatics of the conversation - the social context in which the discussion is undertaken. Several additional models have been proposed as a means to encapsulate the pragmatics of a particular social interaction; however each approach introduces a layer of additional complexity. For example, LAP, or the Language Action Perspective, was proposed as a way to associate language with action and thus provide more effective coordinative systems (Winograd, 1987). However, the approach came under criticism due to the need for users to categorise their messages prior to interaction. Meaning or intention, as argued by Suchman, cannot be categorised prior to social interaction but conversely only emerges through social interaction (Suchman, 1993). A second criticism of LAP and similar systems is their tendency to rigidly define communication into a narrow set of inflexible categories (Ljungberg & Holm, 1996). This can be quite frustrating for users, and is possibly more appropriate when the group is small or select and the task better understood. Other approaches, such as Cope IT (Tzagarakis, Karousos, Karacapilidis, & Nousia, 2009), focus on the core tenets of argumentation and implement formal models similar to Toulmin's categories (Toulmin, 2003). These systems can be beneficial in terms of analysis and decision-making but, again, are inflexible and define communication within a narrow scope. Heer's CommentSpace, in contrast, uses lightweight tagging to provide more structure around the analytic process without introducing the inflexibility of other systems such as Cope IT (Willett & Heer, 2011). Each approach is not without merit however, and the visualisation designer must consider the advantages of introducing more structure with the potential of frustrating users with rigid discussion architectures. Based on this analysis, we draw our final design consideration:

DC30. **Commentary**: Balance flexibility and ease of use with the group size and level of expertise.

3.7.4 Discussion

Although collaborative support may be beyond the requirements for a visualisation designer, implementing simple collaborative mechanisms can increase the potential for users to cohere around visual media. However, providing collaborative support is not always appropriate and the visualisation designer must consider how the group interacts, their size and level of expertise. In terms of evaluation, Name Voyager, Sense.US and Many Eyes are of particular note given how the authors sought to evaluate each system in the context of its intended use. Moreover, Name Voyager and Many Eyes emerged as popular social applications based, in part, on their simplicity and ability to encourage participation. Exploring collaboration requires researchers to develop systems and applications that can be deployed and then studied; however, lab experiments and user studies can be used to focus on particular aspects or techniques of a system or application. CommentSpace, for example, is evaluated in this way; first with a lab study focusing on a given aspect of the tool and then with a live deployment. Table 7 outlines how each design consideration has been evaluated.

Table 7: Evaluation of selected design considerations focused on collaboration

	CodeSaw	Sense.US	Name Voyager	Many Eyes	CommentSpace
In-View Annotation	**PDUS**	**USLD**	Ne	Ne	Ne
Simplicity	**PDUS**	Ne	**LD**	Ne	Ne
Sharing	**P**	**USLD**	**LD**	**LD**	Ne
Personal Relevancy	**PDUS**	Ne	**LD**	Ne	Ne
Commentary	**P**	**USLD**	**LD**	**LD**	**LELD**

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used for academic analysis
Ne - Not Evaluated

3.8 Conclusion and Future Directions

One of the underlying aims of this thesis is to examine how we can develop a more reflective approach to online community design through the application of visual representation. The design considerations that we developed in this chapter are an attempt to identify the principles that underlie the effective visual representation of online communities, not only in terms of graphical presentation but also in terms of how we engage novice or more casual users. The ability to design interesting and provocative visualisations, which address points of concern in intuitive ways, contributes to the success of much of the work in social visualisation and social data analysis. Similarly, the aim of this set of considerations is to support the design of visualisation systems that encourage analysis, discussion and reflection. In Table 8, we summarise how each design consideration has been evaluated in the reviewed literature. We include in this table the two case studies that we conduct in the following chapters 4 and 5.

Table 8: Table of selected design consideration for the visual representation of online communities

	ICTA	Chan & Hayes	Furtado & Andrade	People Garden	Communication Gard	Loom & Loom 2	history flow	Authorlines / NGC	Chromogram	Multivariate Graphs	SaNDVis	Smith et al.	MatrixExplorer	HoneyComb	Visual Who	Bloom	Post History	Viztser	Heer & Roberston
DC1 Network Extract	**LE**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	**LD**	**AT**	Ne	Ne	Ne	Ne	**US**	**US**	Ne
DC2 Interaction	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC3 Contribution	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC4 Playful Mappings	Ne	Ne	Ne	**P**	**LE**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**P**	Ne	**P**	**P**	**US**	**US**	Ne
DC5 Multivariate Viz	Ne	Ne	Ne	**P**	**LE**	Ne	Ne	**US**	Ne	**US**	Ne	**AT**	Ne	Ne	Ne	**P**	Ne	Ne	Ne
DC6 Artistic Rend	Ne	Ne	Ne	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC7 Corpus Viz	Ne	Ne	Ne	Ne	Ne	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC8 Pixel-orientated	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC9 Spatial Layout	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	**LD**	**AT**	Ne	Ne	Ne	Ne	Ne	**US**	Ne
DC10 Search and Disc	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne
DC11 Iconography	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne
DC12 Social Signature	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC13 Socio-temporal	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	**AT**	Ne	**P**	**P**	**P**	**US**	Ne	Ne
DC14 Matrices	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDLE**	**P**	Ne	Ne	Ne	Ne	Ne
DC15 Asym Networks	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDLE**	**P**	Ne	Ne	Ne	Ne	Ne
DC16 Abs Connections	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Na	Ne	**PDLE**	**P**	Ne	Ne	Ne	Ne	Ne
DC17 HO Abstractions	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne
DC18 Inter Animation	Ne	Ne	Ne	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**P**	**P**	**US**	Ne	Ne
DC19 Inter Explore	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne
DC20 Anim Tran	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LE**
DC21 Coordinated Viz	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDLE**	Ne	Ne	Ne	Ne	Ne	Ne
DC22 Linked Viz	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne	Ne	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC23 Embedded Viz	Ne	Ne	Ne	Ne	**LE**	Ne	**AT**	Ne	Ne	Ne	Na	Na	Ne	Ne	Ne	Ne	Ne	Ne	Ne

DC24 Cross Commun	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC25 Small Multiples	Ne	Ne	Ne	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC26 Annotation	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC27 Simplicity	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**US**	Ne
DC28 Sharing	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC29 Personal Rel	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC30 Commentary	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne

P - Prototype only, but not evaluated

LE - Prototype evaluated in lab experiments

US - Prototype evaluated in a user study with authentic users

PD - Participatory design process with likely users

LD - Prototype evaluated in a live deployment or via longitudinal case study

AT - Analytic tool used by academic analysis

Ne - Not Evaluated

Analytic Abstraction

Aesthetic Visual Mappings

Mapping Community Networks

Interaction and Animation

Integration and Configuration

Collaborative Support

	SocialAction	NodeXL	IChase	MacLean and Hangal	CommunityCompare	Furtado & Andrade	Conversation Thumb	CodeSaw	Sense.US	Name Voyager	Many Eyes	CommentSpace	**Case Study 1**	**Case Study 2**
DC1 Network Extract	**LD**	**LD**	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC2 Interaction	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC3 Contribution	Ne	Ne	Ne	Ne	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**
DC4 Playful Mappings	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC5 Multivariate Viz	Ne	Ne	**PDUS**	**AT**	**US**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**
DC6 Artistic Rend	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC7 Corpus Viz	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC8 Pixel-orientated	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC9 Spatial Layout	**LD**	**LD**	Ne	**AT**	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne	Ne	**PDUS**
DC10 Search and Disc	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**
DC11 Iconography	Ne	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**
DC12 Social Signature	Ne	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC13 Socio-temporal	**LD**	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC14 Matrices	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC15 Asym Networks	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC16 Abs Connections	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne

DC17 HO Abstractions	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC18 Inter Animation	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**	**USLD**	**LD**	**LD**	Ne	**LD**	Ne
DC19 Inter Explore	**LD**	**LD**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne	Ne	Ne
DC20 Anim Tran	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**
DC21 Coordinated Viz	**LD**	**LD**	**PDUS**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**LD**	Ne
DC22 Linked Viz	Ne	Ne	Ne	**AT**	Ne	Ne	Ne	Ne	**USLD**	Ne	Ne	Ne	Ne	**PDUS**
DC23 Embedded Viz	Ne	Ne	Ne	Ne	**US**	Ne	**P**	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC24 Cross Commun	Ne	Ne	Ne	**AT**	**US**	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne	Ne
DC25 Small Multiples	Na	**LD**	Ne	Ne	Ne	Ne	**P**	**PDUS**	Ne	Ne	**LD**	Ne	Ne	Ne
DC26 Annotation	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**	**USLD**	Ne	Ne	Ne	**LD**	Ne
DC27 Simplicity	Ne	Ne	Ne	Ne	**US**	Ne	Ne	**PDUS**	Ne	**LD**	**LD**	Ne	**LD**	**PDUS**
DC28 Sharing	Ne	**LD**	Ne	Ne	Ne	Ne	Ne	**P**	**USLD**	**LD**	**LD**	Ne	**LD**	Ne
DC29 Personal Rel	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**PDUS**	**USLD**	**LD**	**LD**	Ne	Ne	Ne
DC30 Commentary	Ne	Ne	Ne	Ne	Ne	Ne	Ne	**P**	**USLD**	**LD**	**LD**	**LELD**	**LD**	Ne

P - Prototype only, but not evaluated
LE - Prototype evaluated in lab experiments
US - Prototype evaluated in a user study with authentic users
PD - Participatory design process with likely users
LD - Prototype evaluated in a live deployment or via longitudinal case study
AT - Analytic tool used by academic analysis
Ne - Not Evaluated

Analytic Abstraction
Aesthetic Visual Mappings
Mapping Community Networks
Interaction and Animation
Integration and Configuration
Collaborative Support

As discussed in the introduction to this chapter, we consider this set of design considerations as an intellectual tool - or design framework - that enables us to address the design of different visualisation systems in the remainder of this thesis. However, we also anticipate other visual designers and visualisation researchers using this framework to address a broad range of design problems, particularly when engaging novice users. In this context, this framework provides us with a research agenda for future research not just into the visual representation of online communities but also in terms of how to support reflective community design. For example, there is considerable scope to examine how collaborative visual analytics can support analysis and discussion in online communities (chapter 4 of this thesis). This has only been touched upon in the literature and requires further empirical investigation. At the same time, however, we require a deeper understanding of how to address the needs of online community designers (chapter 5 of this thesis). Only four of the papers discussed in this chapter, see Table 8, include users in the design process (three approaches include participative design and one approach (CommunityCompare) includes motivating interviews). By implementing this set of design considerations into working systems, we can demonstrate their feasibility and also, as suggested by Hevner et al., evaluate their suitability to the context of use (Hevner et al., 2004). This allows us to consider their advantages and disadvantages but also to suggest improvements based on empirical evidence. This is of particular importance because, although drawn from the literature, a criticism of some of this work, particularly in the social visualisation literature, is the lack of evaluation.

In the following chapters we conduct two independent case studies in which we apply aspects of this design framework and evaluate feasibility of different design elements to the appropriate context. Popular in social sciences, and increasingly used in information systems research, case studies are suited to the analysis of complex systems conducted within the context of their environment (de Moor, 1999; Flyvbjerg, 2011). Online communities are both situated and context-dependent. Our selection of the case study methodology, thus, is based on this reasoning; we require an approach that is suited to developing context-dependent knowledge.

As shown in Table 8, there are different ways in which to implement and evaluate the design framework developed in this chapter. Each community, whether Wiki-base or Q&A for example, will potentially require a different implementation, tailored to their existing goals and practices. However, in this thesis, due to primarily resources, we focus on two specific case-based implementations. Each addresses a particular problem that although suggested in the literature has not been investigated in depth. The first explores how collaborative visual analytics can support analysis and discussion in an instance of an online Q&A community called Super User. Out of the papers reviewed in this chapter, only CodeSaw, by Gilbert and Karahalios, implement collaborative support (Sect. 3.7.1). Their approach, called "spatial messaging", enables users to attach or pin messages to the visualisation state, augmenting analysis and enabling reflection. Building on this idea, and based on a subset of the design framework, we implement and evaluate a visual analytics system called Explore.SU. The evaluation is conducted online over a two week period and members of the Super User online community are invited to participate in the study. The results allow us to assess the feasibility and suitability of the design framework, propose alternative implementations and suggest amendments to the design consideration in Table 8. The second case study investigates how visual analytics can be used to support the reflective design practice of online community managers at Symantec. Again, as with chapter 4, we draw on the design considerations developed in this chapter and implement a system called Petri through several participative design workshops conducted with the community management team. The aim here is to look at reflection in a different light, not as group process conducted occasionally but as a way to improve existing managerial practices. We evaluate the feasibility of the design framework through a grounded evaluation conducted with the community management team at their offices in Dublin.

4

Case Study 1

Reflecting on community Q&A

"Talking with others — that's really another form of reflective conversation. Not only do I think about something differently when I have cast it into a material form, but when I show it to you, you too will think of it differently, and differently from me. Then we talk, argue, joke about or mull over our differences, and as we do so the idea becomes richer."

Thomas Erickson, Socio-Technical Design, 2009

4.1 Introduction

The previous chapter proposed a set of design considerations that can guide the visual representation of online communities. In this chapter, we draw on this framework to implement a collaborative visual analytics site called Explore.SU or Explore Super User. Super User is a community Q&A (Question and Answer) site that addresses technical-support questions on a range of software and hardware-related issues. Explore.SU is designed to encourage discussion, analysis and reflection from within the membership of the Super User community. The site includes a suite of interactive visualisations and a set of collaboration mechanisms that support visual analysis and facilitate group discussion. We deploy the site on the web, invite members of the Super User online community to participate in the study and closely observe usage of the system over several weeks. The observations from the study enable us to evaluate our design decisions in an appropriate context and to suggest possible improvements.

4.1.1 Methodology

The author of this thesis is a member of the Super User community. Drawing on this experience, we analyse different aspects of the community and, based on this analysis, select a subset of design considerations from Table 8. We do not focus on a particular user group, such as community moderators or community elders, but instead focus on how collaborative support can encourage analysis and discussion across the community membership. Having discussed our selection of design considerations, next we describe how the different design elements are implemented in Explore.SU. In addressing the third research objective of this thesis (RO3), we deploy Explore.SU on the web and invite existing members of the Super User community to participate in the study. This enables us to conduct the study with the "correct users" (Andrews, 2006), described by Ellis and Dix as "worth their weight in gold" (Ellis & Dix, 2006), in an “ecologically valid scenario” (Bernstein et al., 2011). Chi argues that addressing communities on the web in this way is an increasingly important practice for the development of social computing research, and, while introducing additional challenges in terms of implementation and scientific rigour, the results can be of more benefit than when a similar study is conducted in the lab (Chi, 2009). In terms of this thesis, this approach allows us to better understand the feasibility of the different design elements than if we are to conduct a similar study under laboratory conditions. Given that participants in the study are novice users, we decide upon an exploratory evaluation methodology that allows for discovery, observation and insight generation without the rigidity of pre-determined tasks evaluated (Saraiya et al., 2005). Finally, given that aim of information visualisation is to generate insight, we use an approach similar to North’s insight-based methodology, to identify and categorise the depth and level of insight generated during the evaluation (Saraiya et al., 2005). Based on this data, coupled with an analysis of exit surveys, we discuss the design considerations, compare and contrast their implementation and suggest possible improvements.

4.1.2 Chapter Outline

The chapter is divided into five sections. First, in Sect. 4.2, there is the analysis and selection of design considerations. Drawing on the author's experience in the Super User community, we analyse different aspects of the community’s purpose, goals and socio-technical architecture and select a set of design considerations that guide the development of a collaborative visual analytics system. Second, in Sect. 4.3, we describe the implementation of the different design elements in Explore.SU. Third, in Sect 4.4, we describe our approach to evaluation and present results. Next, in Sect. 4.5, we discuss some limitations and finally, in Sect. 4.5, we present a discussion in which the feasibility of design approach is considered.

4.2 Analysis and Selection of Design Considerations

Super User[46] is an online Q&A community dedicated to addressing software and hardware technical support questions. At the time of this study, the community had approximately ~50,000 members. As mentioned in Sect. 4.1.1, the author of this thesis is a member of the Super User community and his experience as a community member was used to guide the selection of relevant design considerations. Super User launched in 2009 and has since grown considerably, ranking as one of the largest sites on the Stack Exchange network. Stack Exchange (SE) has made a notable contribution to the design and development of communities on the web, implementing interesting variations on some of ideas that first appeared in sites such as Wikipedia. Ultimately, the aim of the Stack Exchange is to crowd-source Q&A but there is great emphasis placed on building sustainable and productive online communities. The founders, Jeff Attwood and Joel Spolsky, two prominent internet bloggers and entrepreneurs, have written extensively on the subject of online communities and established Stack Overflow[47] - the primary reference for programmers on the web - as an answer to the deficiencies they identified in the existing crop of online Q&A communities. Super User is a very different site than Stack Overflow, the community is much smaller and less active and the aim is to address technical support as opposed to programming-related questions. This way, Stack Overflow remains a place for programming and problem solving, while more support-orientated questions can be directed (or in the parlance of Stack Exchange "migrated") to Super User. Nevertheless, the basic socio-technical configuration remains the same across all Stack Exchange sites. For example, drawing on Wikipedia's model of peer production, Super User is collaboratively edited. Once users have developed enough reputation, they can review each other’s

[46] www.superuser.com

[47] www.stackoverflow.com

posts, an activity that is heavily promoted across all SE communities. Other aspects of the socio-technical configuration are also noteworthy. Comments, or supplementary textual communications appended to posts, enable users to collaborate without inserting irrelevant discussion into the primary text. Similarly, as a member of Stack Exchange, a section of Super User is reserved for the community to consider the design and continued development of their own socio-technical environment. This approach to online community design was first introduced with MetaFilter[48], as discussed in Sect. 2.3.2, and later implemented in Wikipedia as the Village Pump[49].

In the rest of this section we describe the rationale for the decisions that we took in the design of Explore.SU. Each decision is based on the design considerations presented in Table 8. Although the implementation enables us to assess the feasibility of the design approach within context, the underlying aim of the Explore.SU is to catalyse reflection in the community.

4.2.1 Reveal Socio Temporal Dynamics

The aim of our approach was to design visualisations that revealed the social dynamics of the Super User to the community members, or at least those who participate in the study. We did not want to focus on a particular aspect of the community, such as tags or tag co-occurrences, nor did we want to focus on particular threads or subjects within the Super User subject taxonomy. This is because, first, these questions may only hold relevance for a select set of community members and, second, many of these questions can be readily addressed using the Stack Exchange data explorer[50] - a simple interface that enables users to compose queries across each of the communities hosted on Stack Exchange. These queries can then be saved, browsed and executed by other users[51]. Moreover, the aim of the approach was to reveal to the community the underlying social dynamics in a global context. This is a challenging objective, if we consider that the community had over ~50,000 members at the time of the study. From the literature reviewed in chapter 3, social dynamics can be addressed in number of different ways and at different levels of resolution. For example, Smith et al., in Sect. 3.4.2, explore the social dynamics of individual users in Usenet, while Van Ham et al. use Honeycomb to explore the social dynamics of groups within an organisation (Sect. 3.4.3). Alternatively, in Sect. 3.3.3, Víegas et al. show that visualising the evolution of Wikipedia pages is an effective way to consider social dynamics, while iChase, discussed in Sect. 3.6.1., conveyed the social dynamics of Wikipedia by aggregating several sources into a single coordinated visual interface. Considering the scale of the community, and the potential density of the network, we based our approach on several design considerations drawn from Sect. 3.4.3. We used adjacency matrices as opposed to node-link diagrams to represent the network **(DC14)**, asymmetric connections to illustrate the flow of communication **(DC15)**, absent connections to show points of inactivity **(DC16)** and higher-order abstractions to drive analysis **(DC17)**. This final design consideration presents a challenge in online community environments given the absence of a concrete organisational structure to help organise the visual representation. However, as we discuss in the next section, reputation, which is an integral part of the Stack Exchange experience, can be substituted as an organising principle. Finally, social dynamics include both social and temporal structures. According to Fisher and Dourish, social structures describe patterns of social interaction, while temporal structures describe how these patterns change over time (Fisher & Dourish, 2004). Revealing social dynamics can,

---

[48] www.metafilter.com

[49] Wikipedia Village Pump, en.wikipedia.org/wiki/Wikipedia:Village_pump, is described as "a set of pages is used to discuss the technical issues, policies, and operations of Wikipedia".

[50] The data explorer is available at data.stackexchange.com and while this interface is used by those interested in data exploration, there are only 37 queries super User saved at the time of writing.

[51] In 2012, Stack Exchange ran a competition to crowd source interesting data visualisations of the Stack Exchange Network. The results are available here: http://www.kaggle.com/c/predict-closed-questions-on-stack-overflow/prospector#213.

thus, include both a social and temporal dimension. Representing temporal shifts in a network can be addressed in two ways, through the application of interactive animation **(DC18)** or via small multiples **(DC25)**. Each has their advantage, it has been shown that the first provides an engaging experience while the second can improve perception. Given our focus on novice users, we decided to use interactive animation to enable users explore the network over time (**DC18**).

4.2.2 Reputation as an organising principle

Reputation is a core element in how Super User is organised. Stack Exchange was one of the first sites to introduce "Gamification" or, as defined by Deterding et al., "the use of game-like elements in non game contexts" to community Q&A (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). In short, users who ask relevant questions or provide useful answers are automatically awarded reputation points by the system. However, users can also attribute reputation points to other users based on posts they find interesting or particularly relevant. This helps to surface quality content in the system and reduce the signal to noise-to-noise ratio that came to challenge the efficacy of previous systems. Further, reputation reflects the communities trust in a particular user and as the user's reputation increases, they are automatically given more responsibility and privileges by the system. For example, a user with a reputation of 15 is allowed to up-vote a post, while a user with a reputation of 2000 is allowed to edit another user's posts. Badges, see Figure 30 below, are also awarded to users who complete specific milestones.

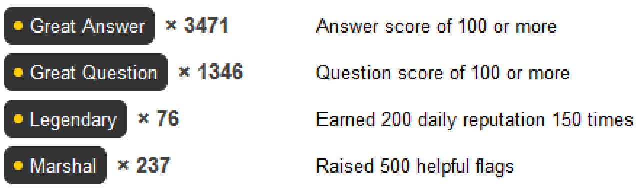


Figure 30: Super User Badges

Community members inherently understand the significance of reputation in Super User. It is one of the driving influences in the Stack Exchange network and is, arguably, one of the reasons for Stack Overflow's success. Reputation provides us with a relevant organisational "abstraction" that can be used to reveal the social dynamics in the Super User community. For example, using reputation as an invariant, and coupling a user's reputation with their connections in the community, we can consider how reputation impacts interaction between community members. If there are dominant flows of communication, for example, or if certain social patterns are more prevalent based on the user's level of reputation.

4.2.3 Simplicity as a design criterion

One of the design considerations that we sought to emphasise with the Explore.SU was simplicity **(DC27)**. Simple design, as discussed in Sect. 3.7.1, is potentially one of the driving reasons for the success of approaches such Name Voyager and Many Eyes. The design is familiar to users, piques their interest, due to the use of personally relevant data in some cases **(DC29)**, but is also intuitive and easy to use. However, designing a simple interface that allows users to explore the dynamics of a large community is notably challenging. As identified Xu et al. with CommunityCompare (Sect. 3.6.3), there is the potential to design a complex visual interface that overwhelms the user or a system with too few capabilities that fails to catch their attention. In the first instance, the cost structure of using the tool can be prohibitive, particularly for novice users, while in the second, users will be unable to generate any significant levels of insight. With Explore.SU, we originally considered using a range of different visual representations, similar to the approach undertaken with Sense.US as discussed in 3.7, but were concerned that the addition of each representation increased the inherent complexity of the system's design and implementation. This is of particular importance considering the evaluation included a live deployment on the web with community members. As a result, we opted for simplicity over complexity and implemented a single visual interface that is based on one visual metaphor as opposed to providing users with different visual representations.

4.2.4 Community design as collective endeavour

Co-creation, discussed in Sect. 2.3.2, is an integral part of the Stack Exchange process. Each Stack Exchange community has an area called the Meta-site that is reserved for the community to consider and engage with the design and continued development of their own socio-technical environment. As mentioned, this idea was first introduced in MetaFilter, was later implemented as the Village Pump in Wikipedia and has since been established as means to facilitate product development across community focused sites. This approach democratises the community design process and enables users with different contributory profiles to participate in community development. Our aim with Explore.SU was to replicate, in a relative way, this process but to include visualisation as way to catalyse analysis and discussion. As shown in the previous chapter, approaches such as NameVoyager, Sense.US and Many Eyes have illustrated the ability for groups to coalesce around the analysis of data, annotating, discussing and sharing visualisations as part of social data analysis. Our aim is to draw on the success of this work and design an approach that encourages analysis and discussion amongst the community. Several design considerations, discussed in Sect. 3.7, address these concerns. For example, considering the group, and based on the our aim of addressing a broad section of the community, we require a simple discussion model **(DC30)** that is familiar to users of the system. Similarly, simple annotation tools can support reference and, potentially through playful commentary, catalyse additional contribution **(DC26)**. Finally, sharing is also required so that users can reference points in the visualisation, enabling grounding and contextual analysis **(DC28)**.

4.3 Implementation of Design Considerations

Having discussed the rationale behind the selection of design considerations, in this section we discuss their implementation in Explore.SU. As mentioned, Explore.SU provides a suite of web-based interactive visualisations that enable users analyse and discuss the behaviour of the Super User online community.

4.3.1 Visualisation

The size of the community (~50,000 users at the time the project was undertaken) made rendering a single one-to-one adjacency matrix unfeasible. Rather, drawing on the design considerations presented in Sect. 3.4.3, we selected reputation as a higher-order abstraction that would help to drive analysis. Although the resolution of analysis is reduced with this method, as the user is unable to locate or select individual community members, the ability to generate insight into the dynamics of community remains supported. However, the distribution of reputation is continuous and based on a log scale. For example, a large number of users had a reputation of below ten (~26,655), while only a relative few had a reputation of over ~50,000. To represent this distribution, and ensure that we design a visual representation that users of the Explore.SU can understand, we took a pragmatic approach and divided reputation into a number bins, plotted on both x and y axes. The first 10 bins are increased by a factor of 10, so 0-10, 10-20, 20-30 etc. The next ten bins are increased by a factor of 100, so 100-200, 200-300, 300-400 etc. The next ten bins are increased by a factor of 1000, so 10000-2000, 2000-3000, 3000-4000 and finally the remaining bins are increased by a factor of 10,000, so 10,000-20,000, 30,000-40,000 etc. Saturation is then used to illustrate the number of connections between bins. This is a pragmatic approach in the sense that we are able to address scale in a relatively straightforward way, providing a visual composition that is both clear and comprehensible, but also reflect the scale-free network topology familiar in large, open web-based systems. We also sought to use asymmetric connections to illustrate the flow of communication in the community **(DC15)** and use absent connections to illustrate points of inactivity **(DC16)**. To this end, we generated a graph using the community's reply-to network. As discussed in 3.2.1, this involves the following principles. If *user A* replies to *user B*, a connection is created between those two users. However, due to the fact that communication in online communities is non-reciprocal, *user A* replying to *user B* creates an in-link between those two users without the need for *user B* to reply-back to *user A*. Plotting the in-links between users creates an asymmetrical representation, which not only highlights areas of activity but also emphasises areas of inactivity via the absence of connections. Experimenting with these design considerations resulted in a compelling way to drive analysis, enabling the identification of points where one section of the community clearly and specifically addresses another. In Figure 31, for example, the distribution for both revisions (image 1) and comments (image 2) is illustrated. Clearly, there is a strong correlation between reputation and revision (illustrated in image 1) while comments (illustrated in image 2) are much more evenly distributed. Users, regardless of reputation, make use of comments as a way to seek and provide clarification as regards the content of their posts. In contrast, revision is generally conducted by users with a high reputation on posts contributed by users with a lower reputation. This is expected as reputation is associated with the user's knowledge of the subject, their awareness of social conventions and their general contribution to the community. The same visualisation is used to provide access to a range of different communication patterns such as posts, comments and accepted answers. Accepted answers are replies that the original poster (or OP) accepts as a solution to their question. This is a particularly important pattern in Super User as it illustrates the sections of the community provide the most value.

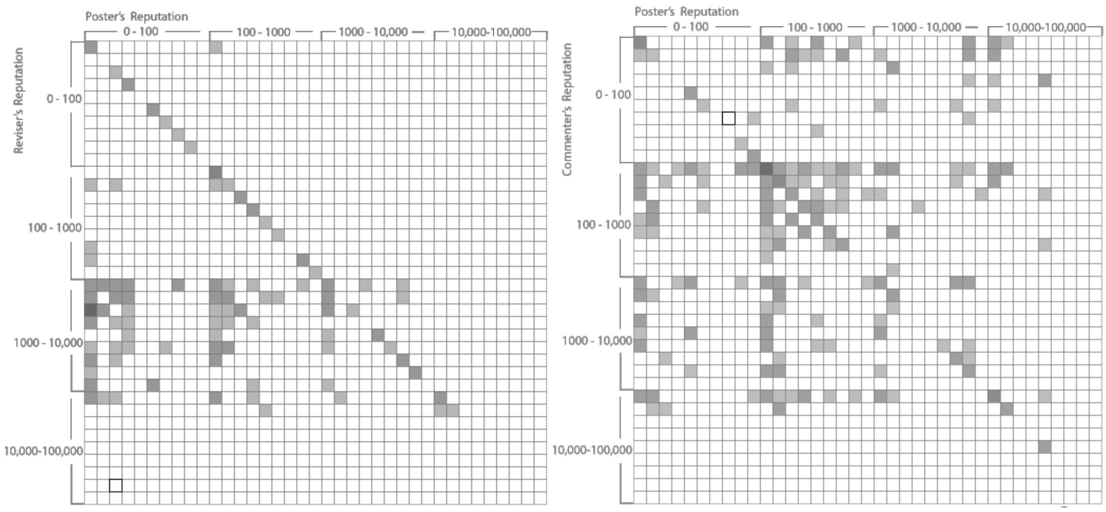


Figure 31: Two images drawn from Explore.SU. The image on the left illustrates the distribution of revisions while the image on the right distribution of comments. There is a clear convention around revision. Users revise their own posts but also users with a high reputation revise other user's posts.

4.3.2 Interaction

To address the temporal dimension of social dynamics, we decided to implement an approach to interactive animation **(DC18)**. This includes an adjustable timeline (as a line graph and slider in Figure 35), which enables users to adjust the date, which in turn updates the adjacency matrix. Dragging the slider continually updates and thus animates the matrix, enabling the user to quickly observe temporal shifts in the community's social patterns. Finally, selecting any glyph in the matrix generates a popup summarising the activity for that glyph (see Figure 32).

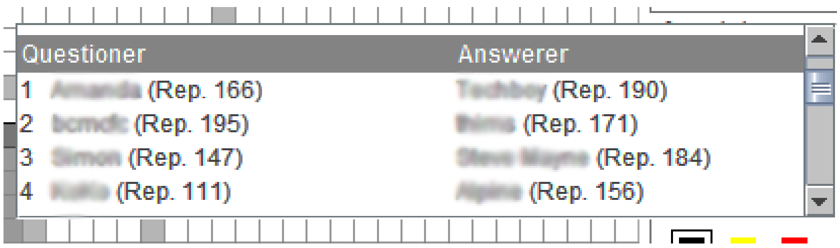


Figure 32: Clicking on any glyph on the matrix will generate a popup, which displays the activity represented by that glyph. In this example, the popup displays questions and answers from users with a reputation of 100 to 200 giving a global and local representation of the community's activity.

4.3.3 Collaboration

One of the underlying aims of Explore.SU is to catalyse analysis and discussion with members of the Super User community. From the discussion in Sect 4.2, we required a way for users to highlight points of interest, share their observations with other users and discuss these observations in a simple yet engaging manner. Drawing on the design considerations presented in Table 8, we implemented several collaborative mechanisms, including in-view annotation **(DC26)**, sharing **(DC28)** and commentary **(DC30)**. It is important to note that few of the approaches considered in Table 8 have addressed the design of collaborative support at this level of detail and the approaches that include some form of collaborative support have limited evaluations. Thus, while collaboration is sometimes suggested by authors, or implemented after an evaluation, the design of collaborative mechanisms have yet to be evaluated with online community members in a live deployment.

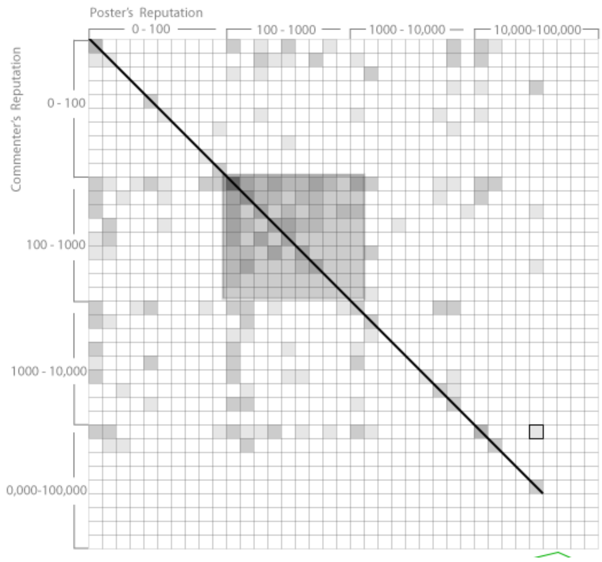


Figure 33: An annotated visualisation from Explore.SU. In this image, a user has drawn a diagonal line to emphasise how users revise their own posts. Similarly, the dark square illustrates an area where activity is most concentrated.

4.3.3.1 In-view Annotation

We implemented several graphical annotation tools to help users contextualise their observations. Pointing is the most important tool, providing users with a way to gesture towards an object of interest. This is implemented using a simple tool that enables the user to draw an arrow on the visualisation as if drawing on an acetate sheet. However, users were also able to emphasise and highlight different aspects of the visualisation by using different geometrical shapes, such as squares and lines as illustrated in Figure 33. Each annotation is associated with the visualisation's view state.

4.3.3.2 Sharing

To address sharing, we implement simple collaborative functionality in the form of visualisation bookmarking and view sharing. Again, drawing from the Sense.US as discussed in Sect. 3.7.2, bookmarking enables users to capture the state of a visualisation as visualisation view (a view is a small png that includes any user annotations) and then seamlessly share that view as an attachment to a comment (Figure 34). Clicking on a view appended to a comment then enables the user to load that visualisation state, coupled with any graphical annotations, into the main visualisation area. The user can then contribute additional annotations and then save that new visualisation view as an attachment to a new comment.

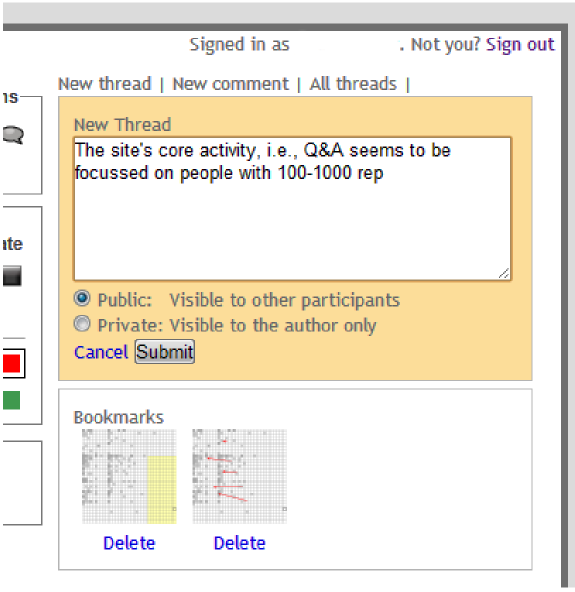


Figure 34: A comment with two attached bookmarks. Each bookmark has annotations added by a user. Clicking the bookmark will load the visualisation, coupled with the annotations, into the interactive Java applet.

4.3.3.3 Commentary

In terms of discussion, we aimed to implement a simple model that did confuse or exclude participants from contribution. Based on an analysis of the community, and considering the discussion in Sect. 3.7.3, we implemented threaded commentary enabling users to elaborate on their annotations, describe observations in more detail and share these observations with other participants in the study. Users can reply to other comments in the thread, providing a space for collaborative analysis. As illustrated in Figure 33, the visualisation state can be bookmarked as a view attached to a comment.

4.3.4 Integration and Configuration

We integrated each of the different design elements into a single interface, illustrated in Figure 35. The interface is divided into four main sections. Section A, is a matrix visualisation of the community's reply-to graph (as discussed in Sect. 4.3.1). Section B is an interactive timeline that enables users traverse time and update the matrix to a given period. Section C is a set of graphical annotation tools and bookmark functionality, discussed in Sect. 4.3.3, which allows users to annotate the visualisation, bookmark annotations and share those bookmarks with other users. Section D provides threaded commentary, allowing users to participate in discussion as regards the different observations drawn from the visualisation.

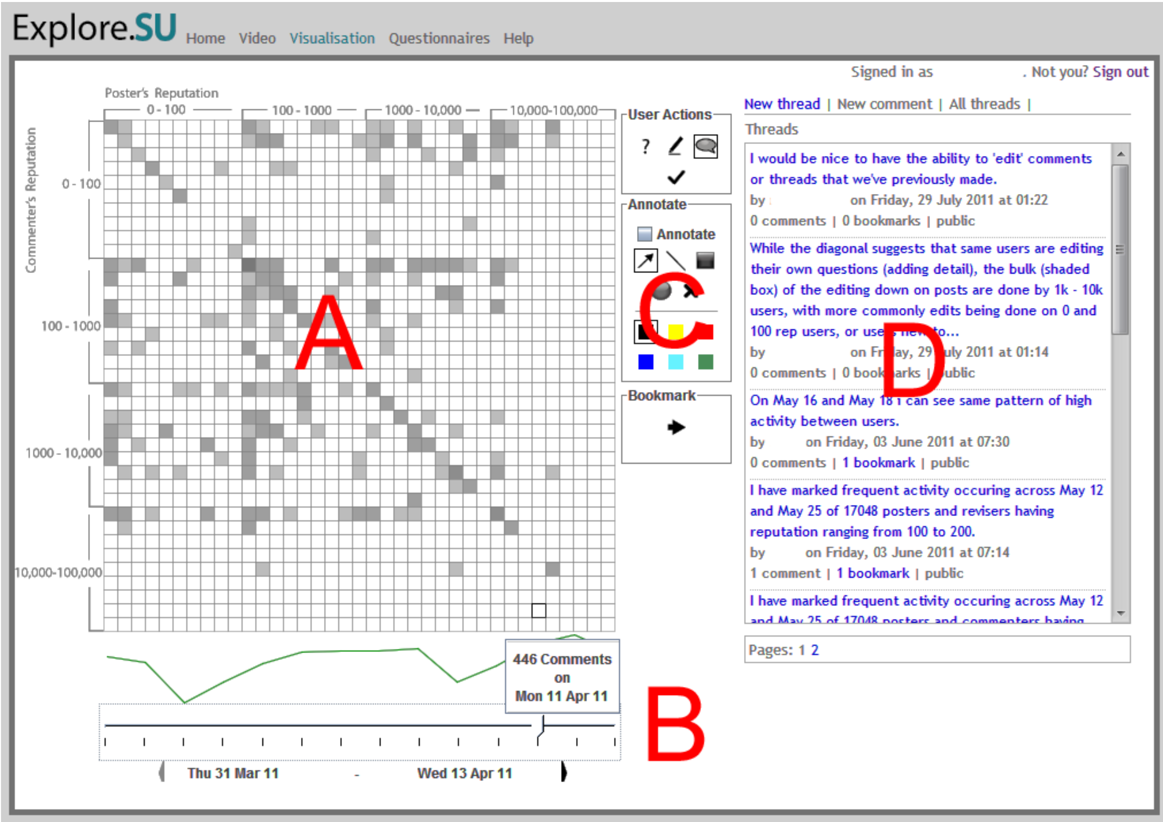


Figure 35: The Explore.SU visual analytics system. The interface is comprised of the following components: A) An interactive Java Applet with graphical annotation functionality. B) Interactive Timeline (graph and slider) enabling the traversal of time. The green line graph illustrates variations in community activity. In the above illustration, each dip in green line graph illustrates a dip in activity at the weekend as the majority of activity on Super User takes place between Monday and Friday or when people are at work. C) User Actions, Graphical Annotation Tools and Bookmark functionality. D) Threaded discussion – a set of threaded comments with visualisation bookmarks that allow users to load a visualisation's state (including any contributed annotations) into the main visualisation area.

4.3.5 Implementation Details

The visualisation was implemented using Java2D as an Applet[52]. The Ruby on Rails (version 2.3) web framework was used to provide server-side support and a MySQL database[53] was used for persistence. The comment functionality (animation and DOM manipulation) was implemented using the Scripticulous[54] JavaScript framework and the Prototype[55] Javascript framework provided AJAX support. We used JSOject[56] as a Java to JavaScript Bridge to enable communication between the java applet and the comment functionality. This was required to create a seamless experience when bookmarking an annotated a visualisation. We implemented social logins using OAuth[57] with the

[52] Interestingly, the use of Java caused issue with some of the participants during the live deployment. Several mentioned a dislike of the language and did not have a JVM (Java Virtual Machine) installed on their machine. While this certainly reduced the number of participants in the study, it is more of a concern that is specific to the community in question. Super User is a technical community and participants indicated a preference for a more lightweight (that does not have a plug-in architecture) implementation such as JavaScript. However, at the time, the author of this thesis had little experience with JavaScript, and not foreseeing this as an issue, chose Java2D to build the visualisation.

[53] dev.mysql.com

[54] script.aculo.us

[55] prototypejs.org

[56] http://docs.oracle.com/javase/7/docs/technotes/guides/plugin/developer_guide/java_js.html

[57] http://oauth.net/

major providers such as Google, Facebook and Twitter. These providers were chosen because, at the time of implementation, Stack Exchange did not provide OAuth authentication. We also provided the ability for users to create their own accounts with Explore.SU.

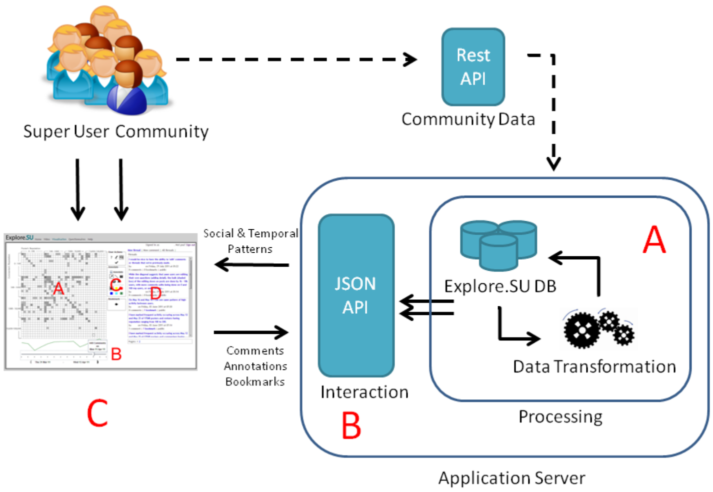


Figure 36: Explore.SU component architecture

Figure 36 illustrates the Explore.SU component architecture. Initially, we imported the monthly data dump[58] provided by the Stack Exchange Network and then synced the local and remote copies of the database every night using the Stack Exchange API[59]. The system architecture is comprised of three principle components. First is the processing component, A in Figure 36, in which the data is transformed. Second is the interaction component, B in Figure 36, which is a JSON RESTful API that manages the interaction between the Explore.SU database and the Explore.SU visualisation interface. Finally, there is a visualisation component, C in Figure 36, which provides visual access to the data and enables users comment, annotate and bookmark the different visualisation views.

4.4 Evaluation of Design Considerations

By implementing the design considerations presented in Table 8 as design elements in the a system and then evaluating that system, we are able to assess the applicability of each design decision in an appropriate context with a representative sample of users. Domain experts, as suggested in Sect. 4.1.1, can hold advantages over subjects with little domain knowledge when pursuing an explorative evaluation of a visual analytics system. Their ability to draw on background knowledge, connect patterns with observations and discover new and interesting insights is clearly of benefit, however, they can also be inherently interested in exploring their domain using new and interesting tools. Furthermore, approaches such as Sense.US and Many Eyes have shown that asynchronous collaborative visualisation can be successfully evaluated in real world scenarios. In terms of Explore.SU, we conduct the evaluation in two stages, a pilot study, which allowed us to refine the prototype and ensure that the user interface was coherent and comprehensible, and a live deployment, which allowed us to observe usage of the system over several weeks. In this section, we describe the procedure (Sect. 4.4.1) and report the results (sect. 4.4.2).

[58] blog.stackoverflow.com/category/cc-wiki-dump/

[59] api.stackoverflow.com

4.4.1 Procedure

In this section, we describe the procedure undertaken in both stages - the pilot study followed by the live deployment. We also briefly describe feedback from the pilot study that we used to refine elements of our system before we deployed Explore.SU on the web.

4.4.1.1 Pilot Lab Study

The pilot study involved seven participants (six males and one female). All participants are computer science PhD researchers from our faculty. None of the participants are online community researchers and their specialities range from semantic technologies to ubicomp systems. Although familiar with the core tenets of information visualisation, none of the participants were visualisation researchers nor were they familiar with collaborative visualisation. In addition, none of the participants were familiar with the Super User online community, however, four had used Stack Overflow previously, and all of the participants were familiar with Yahoo Answers[60] (an alternative non-domain specific question/answer site). The first 5 participants were given a brief tutorial of the system, asked to familiarise themselves with the different interactive elements and then asked to comment on their findings. Feedback suggested the need for a video tutorial, so the final 2 participants were given a video tutorial, outlining the system and the process of the study, asked to familiarise themselves with the system and then comment on their findings. We did not assign any tasks, instead we suggested that participants explore the visualisation, identify patterns of interest and comment on those patterns using the in-view annotation and comment functionality. This is in keeping with the exploratory evaluation methodology mentioned in Sect. 4.1.1 in which the participant is given scope to explore the visualisation without having to complete predetermined tasks. Each session lasted between 50 and 60 minutes, was recorded, notes were taken by an observer and the participants were asked to think aloud.

Feedback from the Pilot Study

Through feedback participants indicated that the interface was coherent and comprehensible. Some expressed frustration at not being able to apply finer grained filtering to the visualisation, to drill down or to compare and contrast the different communication patterns – such as following a specific thread over time or identifying the actions taken during the lifecycle of a thread. Initially, some found it difficult to understand the matrices, especially considering the use of reputation as an organising principle. Only a few users were familiar with the use matrices, most indicating familiarity with node-link diagrams. However, once they grasped the general idea, and understood the visualisation, most users set about identifying patterns and points of interest.

4.4.1.2 Live Deployment

We made some amendments to address the feedback from the pilot study. We added a help section, which described the visualisation and collaboration tools in detail, and provided better labelling to identify the structure and outline of the visualisation. We did not address the suggestions on associating the visualisation to the underlying discussion, even though this is a design consideration **(DC23)**, nor did we add additional interactivity because, as discussed in Sect. 4.2.3, we did not want to increase the complexity of the design. For the live deployment, we recruited members of the Super User online community. A proposal for the study was posted on the Super User meta-site[61]. This was initially endorsed by a community moderator but subsequently removed within twelve hours by a Stack Exchange moderator who argued that this was an advertisement and against Stack Exchange policy. As a result, each day we advertised in the Super User chat room, which tend to be used by the more committed user-base. 28 people signed into the system, 17 completed the evaluation and 56 comments were submitted as part of the study. As with the pilot

[60] answers.yahoo.com
[61] meta.superuser.com

study, participants were asked to sign in, watch a short tutorial video, submit a minimum of five comments and then complete two short questionnaires. The first survey related to the process of the study, inquiring about their experience with the system and if they would be amenable to further contact (Appendix B), while the second survey was a SUS (Simple Usability Score) test that enabled us to assess usability (Appendix A). They were informed that the process would take a minimum of 30 minutes. Six Amazon vouchers were raffled to encourage participation.

4.4.2 Results

Although asked to use the comment functionality provided with Explore.SU, several participants discussed the system in the Super User chat room or posted comments to the Super User meta-site. Some of these comments are included here as they provide some insight into the how users initially perceived and interacted with Explore.SU. Some of the initial feedback was very positive. For example, on the Super User meta-site, one user commented "considering how insightful this is, maybe SE (Stack Exchange) should think of implementing?" Other users took screenshots of the different visual representations and then posted those screenshots to the Super User chat room. For example, one user posted a visualisation of answer activity along with the comment "fascinating, 100-rep people are the ones who're asking the most, 1000-3000 rep ones are the ones to answer the most". This comment included an annotation that referenced the visible pattern and was added using the Explore.SU system. In this case, the visualisation prompted an instance of analysis that the user then shared with the rest of their peers. However, a screenshot of the visualisation was included to lend support to their observation. Later, the same author posted another screenshot from the system and commented "comments and revisions have the same pattern of distribution, obviously comments distribution is more dense [sic]". While the different visual representations enabled the participants to consider the community from different perspectives, their inclusion in the chat room illustrates the potential for visualisation, when drawn into the existing social space, to encourage analysis which in turn engenders further discussion **(DC28)**. Although a single example, these actions was not prompted by the study but were based on the user's interest in sharing their observations with other community members.

In the rest of this section, we report the results of the study. Sect. 4.4.2.1 presents the results of a content analysis that conducted on the comments that were submitted to the Explore.SU system. Sect. 4.4.2.2 describes how the different collaboration mechanisms were utilised. And finally, in Sect. 4.4.2.3, we describe the results of an exit survey and SUS test used to assess experience and usability.

4.4.2.1 Content Analysis of Comments

We conducted a content analysis on the comments submitted to the Explore.SU system. This enables us to better understand how the different design elements contributed to the participant's observations. We also wanted to determine the impact of collaboration on analysis, assessing, for instance, whether an observation developed into a discussion between participants in the study and how the different design elements facilitated this analysis. As discussed in Sect. 4.1.1, we adopt an insight-based methodology to enable us organise and analyse our data. An increasingly popular way to evaluate information visualisation systems (Marie, Card, & Mackinlay, 2009; Ribarsky, 2009), the aim of an insight-based methodology is not to control the evaluation but to enable participants explore the visualisation and report on what they see (North, 2006). Using content analysis, we are then able to code these observations and determine the level and depth of insight generated using the tool. Previous work by Heer at al. illustrates the success of this approach when studying the usage of a collaborative visualisation system called Sense.US (Jeffery Heer et al., 2009). Given the focus of that work on collaborative support, we adopted the coding scheme that the authors developed (inductively) during that study. This coding scheme has the following categories: Observation, Question, Hypothesis, Data Integrity, Linking, Socialising, System Design, Testing, Tips, To-Do and Affirmation. However, five out of the eleven categories proved redundant during coding. The redundant categories were Linking, Socialising, Testing, Tips and To-do. The resulting categories used during coding were Data Integrity, Question, Affirmation, Hypothesis, System Design and Observation. Two individuals coded the comments and we computed the Cohen kappa statistic for inter-rater reliability at 0.77.

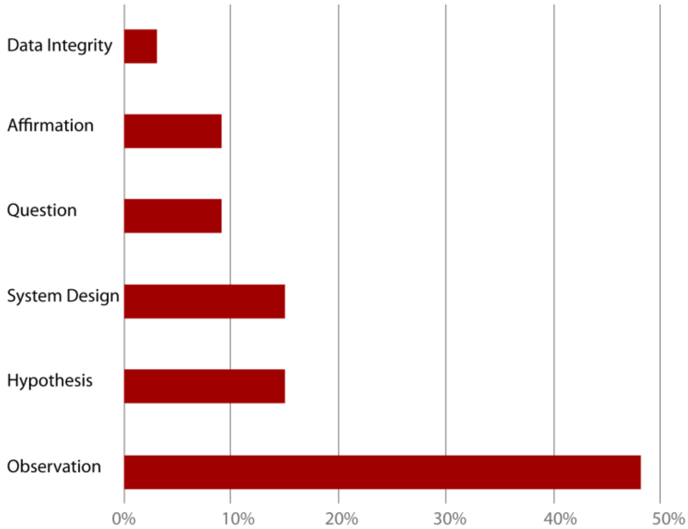


Figure 37: Results of the content analysis on the commentary drawn from Explore.SU. Each category is considered mutually exclusive. As expected, observations were most prevalent category (48%), however, several participants submitted hypotheses (15%) or questioned a previous user's observation (9%).

As illustrated in Figure 37, the majority of comments are observational (48%). Observational comments are simple descriptions of the activity portrayed in the visualisation. For example, the comment, "the site's core activity, i.e., Q&A seems to be focussed on people with 100-1000 rep" is an observational comment in which the participant describes the patterns they observe when using the system. Hypothesis-driven (15%) comments, on the other hand, are more complex, can be supported by existing domain knowledge or result in a question or hypothesis (J.S. Yi et al., 2008). For example, the comment "I would guess that is because the 0-50 and the 100-150 rep ranges represent most new users (either completely new to the SE site, or new to SU)", is an hypothesis-driven comment as the participant draws on their existing knowledge of the community to reason about the patterns that they see when exploring the visualisation. Hypothesis formation is an important step in data exploration, which can lead to further cycles of analysis and discussion. In some cases, participants ask questions of the data (9%), for instance, "Its [sic] also slightly surprising that there aren't more revisions of [sic] by the Poster (diagonal line?)" or affirm an observation submitted previously by another participant (9%), for example, "I noticed that too, and it might be interesting to note that the 100 rep range *can* be users that have migrated from other sites". System design was also drew attention, which included feature suggestions and technical requests (15%), such as, "It would be nice to have the ability to 'edit' comments or threads that we've previously made". While participants used the threaded discussion, just under half the comments had annotations attached (45%) and 15% of comments were submitted as private and thus not available to other participants.

The most interesting instances of analysis came from discussion between participants. For example, in one discussion, several participants considered the community's revision process (illustrated in Figure 31). For example, one participant noted:

> "Posters with a reputation below 500 are often edited, mostly by users with rep (reputation) from 700 up to 10k. People with lower rep (reputation) don't really seem to edit that much, although they theoretically can..." [sic]

While a second replied:

> "The users who appear to be editing posts at lower rep levels appear to be the original question asker. I would guess that at lower rep levels people are much more adverse to 'messing up' other peoples questions and instead focus on their own..." [sic]

Another instance of analysis occurred in a discussion on the relationship between reputation and poster activity. One participant commented that:

> "Although it shifts from day to day, the highest correlation can be found here, either showing that there are many users with rep from 100 to 1k, or they are the ones who answer most, and most questions are posted by users below 1k"

However, a second participant replied:

> "I noticed that too, and it might be interesting to note that the 100 rep (reputation) range can be users that have migrated from other (Stack Exchange) sites."

In this example, the second participant draws on their background knowledge to suggest an alternative hypothesis for why the majority of activity occurs between users with a reputation of around 100. This view is affirmed in a further comment submitted by another participant. Here the participant suggests that the migrated users are well accustomed to the conventions of the Stack Exchange network:

> "I find it interesting that the (100,100) box has had a high number consistently. This furthers xxxx's idea of a poster commenting on his/her own post as the typical 'migrant' from another SE site already knows how to use the Questions and Comments correctly..."

This dominant pattern is the subject of further discussion on another thread. Again, the first participant queries the pattern, while the second suggests that the pattern is due to the Stack Exchange's policy of migrating users between different communities.

> "These two clusters, based on the Questioner's rep, make me wonder how users cross the gap from 50 to 100 rep. It's also interesting that by far the most activity seems to be around these two clusters and not on higher Questioner's rep levels."

While the second participant comments:

> "I would guess that is because the 0-50 and the 100-150 rep ranges represent most new users (either completely new to the SE (Stack Exchange) site, or new to SU (Super User))."

In these instances, the visualisation enables participants to identify several patterns in their community and to attribute meaning to those observations based on their understanding of existing community practice. As illustrated, collaboration plays an important role in this process by enabling the collaborative development of insight and by providing a shared context for reflection.

4.4.2.2 Usage of In-view Annotation Tools

We also wanted to understand how the in-view annotation functionality contributed to discussion and analysis. If, for instance, a certain tool was used more consistently, was there evidence of a preference or a combination of different tools. Annotation proved particularly noteworthy and 45% of comments had an annotated visualisation included as an attachment. Annotations were divided between arrows, used by participants to identify points of interest (28.0%) and other shapes (72.0%) - lines (8.0%), rectangles (62.0%) and circles (2.0%) - used to highlight, fence-off or isolate areas of interest in the visualisation. After arrows, rectangles were the most popular means of annotation. However, some participants used lines to draw triangles to bound an area of interest (see Figure 38 for a summary of usage).

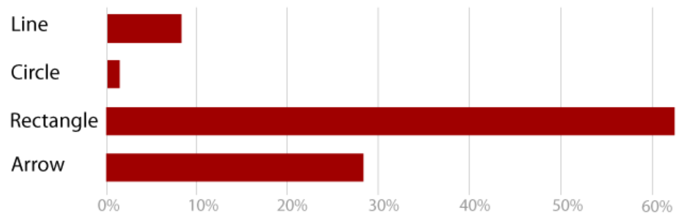


Figure 38: Usage of graphical annotation functionality

4.4.2.3 Usability and Experience

We also asked participants to complete a short exit survey and SUS (Simple Usability Scale) test (Brooke, 1996) designed to gauge their experience of using the Explore.SU system[62]. Overall, participants perceived the system as relatively straightforward and easy to use (based on SUS score of 70.29). In the exit questionnaire, participants responded positively (with an average rating of 4.5/5.0) when asked "Did you enjoy using Explore.SU?" Similarly, when asked "Did you learn something about the Super User community when using Explore.SU?" participants responded with an average rating of 3.8/5.0. Some participants left additional comments to support their answer. One participant, for example, mentioned that Explore.SU "gave me good insight into the dynamics of the community" while another suggested "it was a very interesting and innovative way to view the available data". One participant emphasised the importance of collaboration on their experience of using the system, "I like how you can analyze this in a social way, reading what others have discovered". Finally, two participants, one of whom is a community moderator, remarked on how the visualisation made them reflect on the community's development. One participant commented, "Its [sic] interesting to see what users make up the bulk of the activity, because it means we should try to do more to appeal to these users" while another noted "I think it's a good idea to get an insight on how the community behaves, allowing us to see where we have a lack [sic] and for example how we could inspire more users of a certain reputation range to do something like perhaps making them feel more welcome..."

4.5 Limitations

Although we consider the study provided some interesting insights in to design visualisations that support reflection in online communities, there are some limitations to the study. The sample size is relatively modest when compared with the population of the community. 28 users signed into the system and 17 completed the evaluation. However, Super User had, at the time the project was undertaken, a population of over ~50,000. As discussed, we ran the live deployment over a two week period and, due to the community's policy in regards to advertising on their site, experienced difficulty in attracting participants. We stipulated that the trial was to run for two weeks and, as a

[62] In keeping with the ethics guidelines of Trinity College, school of statistics and computing, each of the two surveys and all questions in the surveys were optional. However, we explicitly asked participants to answer all questions and complete the SUS survey.

consequence, did not extend the trial once the Stack Exchange moderators had re-posted the advertisement for the study. This problem was difficult to foresee as we had cleared the advertisement with the community moderators beforehand, however, we had not specifically consulted with the Stack Exchange moderators (as this was not considered necessary by the community moderators). While these difficulties impacted on our sample size, and illustrate the challenge of conducting usage studies on the web, it could be suggested that only a relatively small proportion of the community would be interested in conducting such a study. This is borne out by the comment from one of the participants who suggested that, as far as he could gather, anyone interested in participating had already done so. In addition, our reliance on Java 2D package to develop the visualisation also reduced our sample size, mainly because this is a technical community who dislike installing the JDK browser plug-in required to run applet software. Several participants who expressed an interest in the study withdrew because of our use of Java. This, again, was unforeseen, but illustrated the general trend towards in-browser languages that run without the addition of plug-in architectures and influenced our approach to implementation in the following case study (chapter 5).

4.6 Discussion of Design Considerations

In general, Explore.SU provided us with some interesting observations. Notable contributions involved participants debating different community conventions using points drawn from the visualisation coupled with their own background knowledge of Super User. In several instances, participants took contrary positions and used different design elements, such as in-view annotation and sharing, to support their positions, and cycles of analysis and discussion lent support to the feasibility of different design elements. The aim of this section is to analyse the applicability of the design approach in light of this evaluation. We do not seek to generalise beyond the specifics of this case study, as this would require a collection of experiences from a larger number of case studies or a longitudinal case study with a specific community, but rather aim to show how designers can apply and potentially refine this framework within a specific context. In each instance, we suggest amendments to different design considerations in Table 8.

4.6.1 Visualisation

The approach to visualisation was influenced by the following design considerations, namely, using matrices to represent the network **(DC14)**, asymmetric connections to illustrate the flow of communication **(DC15)**, absent connections to show areas of inactivity **(DC16)** and higher-order abstractions to drive analysis **(DC17)**. Using reputation as an organising principle enabled the representation of a large dataset in a compact space, while coupling asymmetric connections with absent connections revealed to participants the inherent social hierarchy that exists in their community. For example, one participant commented "fascinating, 100-rep people are the ones who're asking the most, 1000-3000 rep ones are the ones to answer the most", while others responded about how the asymmetric connections reflected the conventions of their community. In particular, this is the case with revision, which is an important structural component in the Stack Exchange architecture, but was also evident in the other communications such as comments and accepted answers. Participants found little difficulty in understanding the visual representation with one participant in the exit survey describing the approach as an “innovative” way to visualise community activity. This is important given the increasing scale of online community datasets and the limited ways in which large networks can be addressed using visualisation technology. Recognisable higher-order abstractions provide an alternative to the application of ego-centric visualisations implemented with interactive exploration **(DC19)** or search and discover interfaces with visualisation views **(DC10)**. This also holds for the use of matrices as opposed to node-link diagrams. Participants made no negative comments about the use of matrices but, judging from the observations submitted during the study and from the exit surveys, felt relatively comfortable with the approach. Focusing on the use of higher-order abstractions to drive analysis, we suggest the following amendment to **DC17**:

Table 9: Suggested amendment to Higher-order Abstractions

DC17 Higher-order Abstractions	Consider conventions embedded in community practice as higher-order abstractions.

4.6.2 Interaction

As discussed in the previous chapter **(DC18)**, interactive animation provides users with an engaging way to explore community datasets by enabling users observe the ebb-and-flow of community life. The implementation of this design consideration in Explore.SU enabled the comparison of different time frames - by addressing one date and then the next - but was less successful at illustrating the difference between those time frames in terms of evolving dynamics. To show diffusion, for example, there is a need to show the sum and not the difference between time frames. An alternative approach, which may have yielded more success in terms of generating insight, could have involved the implementation of small multiples in which several small visualisations are shown in unison and are indexed by time **(DC25)**. Another drawback, which is evinced in the comments during the pilot study and also, to some degree, was the ability for participants to drill-down into the data, an interactive technique commonly found in coordinated visualisation **(DC21)**. Providing a much more interactive experience, in which pan and zoom supports drill-down and filters enable selection yet animation is used to maintain orientation **(DC18)**, could lead to the generation of more compelling observations while also engaging users with a highly interactive experience. However, this approach would also have increased the complexity of the visual interface. Having said this, there is scope to investigate a highly interactive yet compact matrix visualisation that supports re-ordering and re-encoding yet enables the analysis of networks using higher-order abstractions[63]. This widget could be embedded in a community interface **(DC23)** enabling users to drill into community dynamics and verify observations with access to the conversations and user profiles. However, in terms of Table 8, we suggest the following amendment to **DC18:**

Table 10: Suggested amendment to Interactive Animation

DC18 Interactive Animation	Consider interactive animation carefully if supporting visual analysis.

4.6.3 Collaboration

Our approach focused heavily on collaboration as we sought to determine the extent to which our design decisions encouraged analysis and discussion by participants in the study. Both in-view annotation **(DC26)** and sharing **(DC28)** were implemented and provided users with a more social way to reflect on community dynamics. For example, from the outset of the study, participants took screenshots of the visualisation and then posted these screenshots to the Super User chat room. In one instance, a user annotated the visualisation and then posted that annotated view to the chat room illustrating the utility of annotation but also suggesting that in-application commentary, as we implemented in Explore.SU, is less useful when the community has an existing conversational space. It is, however, clear that enabling a way for users to share visualisations beyond the restrictions of the Explore.SU system may yield additional benefit. This is an interesting avenue for further research, particularly considering that there is already an

[63] Henry and Van Ham have both addressed matrices using a variety of techniques to enable the multi-level analysis of network data. However, each approach, which is considered in chapter 3, is based on designing analytic environments that address expert analysts in various forms.

area of Super User reserved for community-wide discussion. In terms of annotation, 45% of comments included an annotated visualisation. This is one of the most popular features of the application, with several participants highlighting that they used this feature to "socially" to see what other participants had discovered. This illustrates the potential for commentary to lead to additional analysis and reflection. There is scope to investigate this further by wrapping commentary and annotation into a single embeddable widget, which can be used on any JavaScript enabled visualisation and is sharable via embeddable URLs across a community infrastructure. This approach reduces the need to confine the users to a system like Explore.SU but also extends the size of the potential participant audience. Drawing from this analysis, we suggest the following amendment to **DC16**:

Table 11: Suggested amendments to Commentary

DC16 Commentary	Consider comments as in-view annotations or enable discussion within existing social spaces.

4.7 Summary

This chapter investigated how collaborative visual analytics could catalyse analysis and discussion in an instance of a Q&A community called Super User. We implemented a site called Explore.SU, invited members of the community to participate in the study and closely observed usage of the system in order to better understand the applicability of our design decisions. Participants found interest in exploring the visualisation and presented several insights for discussion, using collaborative mechanisms to assist their analysis. At several points comments developed as links in a chain of reasoning between different users based on observations drawn from the visualisation coupled with their existing background knowledge of community conventions. In some instances, contradictions emerged as participant's claims were rebutted or countered with further insights drawn from the system. The live deployment enabled us to consider the success of the different design elements and to suggest alternative implementations and refined considerations. The next chapter presents a second case study in which we investigate how to support the reflective practice of online community mangers at Symantec.

5

Case Study 2
Reflective Community Design at Symantec

Reflective practice is "a dialogue of thinking and doing through which I become more skilful."

Donald A. Schön, The Reflective Practitioner, 1984

5.1 Introduction

Although we has argued that visual analysis can support reflective community design, and have proposed a design framework that can facilitate the creation of effective visual representations, the question remains of how to design visual analytic systems that support the existing practice of online community designers. Only four of the approaches presented in Table 8 include community members or other representative users during the design process. Thus, this chapter, in consultation with the community management team at Symantec, investigates how visual analytics could potentially assist in their community management strategy. We conduct a series of participative design workshops in which we draft a set of design requirements for a visual analytics system. Next, drawing on the design considerations presented in Table 8, we design and implement Petri - an interactive visual analytics environment that enables the analysis of online communities from multiple perspectives. Finally, through explorative user studies, we evaluate the design of Petri with members of the community management team. As in the previous chapter, we consider the success of the different design approach and compare and contrast alternative implementations.

5.1.1 Methodology

Taking a human-centred approach to design requires the designer to understand the problems and requirements of the target audience. In this context we draw on Munzner's model for visualisation design and validation as a way to bracket the design process into a series of interrelated stages, which includes domain analysis as part of the design process (Munzner, 2009b). The domain analysis is of particular importance in the context of this work and we conduct several participative design workshops to understand the practices of online community management and to help identify a set of user requirements. While Munzner argues for validation over evaluation, and we sought to validate the outcome of each stage with the community management team, we conduct an evaluation at the conclusion of the project to help ascertain the suitability of our design decisions. In a bid to maintain consistency across evaluations, and considering that the community management team are novice users with little experience of visual analytics, we take a similar approach to evaluation as we did in chapter 4. We conduct explorative user studies with several members of the team, video each study, transcribe the video and use content analysis to evaluate the depth and level of insight. As discussed previously, we draw on several sources to support this methodology. Ellis and Dix, for instance, argue that explorative user studies conducted with domain experts are more suitable to the evaluation of information visualisation than performance-based strategies (Ellis & Dix, 2006). Andrew's reiterates this point, focusing on the need to address "the correct users" with stable systems (Andrews, 2006). Similarly, Saraiya et al. argue that performance-based approaches narrow the generation of insight to within a set of predefined tasks and reduce the potential for explorative insights (Saraiya et al., 2005). However, having physical access to the participants also enables us to conduct semi-structured post evaluation interviews. Mazza and Berré argue that interviews have the advantage of detecting user opinion as to the usefulness of a particular visual representation (Riccardo Mazza & Berré, 2007). Isenberg et al. describe this as a grounded evaluation strategy, the aim being to evaluate the system in the "context of its intended use" and to use qualitative insights drawn from interviews to help contextualise our observations (Isenberg et al., 2008).

5.1.2 Chapter outline

The chapter is divided into five sections. First, in Sect. 5.2, there is the analysis and selection of design considerations. We present findings from several participative design workshops, conducted with the community management team, in which we elicit a set of user requirements for a visual analytics system. Second, in Sect. 5.3, we draw on the design considerations presented in Table 8 to implement a visual analytics system called Petri. Third, in Sect 5.4, we describe the approach to evaluation and present results. Next, in Sect. 5.5, we discuss some limitations. Finally, in Sect. 5.6, we discuss the applicability of our design decisions within the context of this application.

5.2 Analysis and Selection of Design Considerations

In this section we first describe the online community, consider the purpose and general socio-technical configuration employed by Symantec (Sect. 5.2.1). Next, as part of a domain analysis, we report on three workshops that were held with the community management team (Sect. 5.2.2). The aim of the first workshop was to develop a better understanding of the practice of online community management, as a set of operational and strategic directives, and to understand the requirements that the team had in terms of visual analytics. The second and third workshops were used to gather feedback on initial designs.

5.2.1 The Community

The community is a non-English speaking Norton-support forum hosted by Symantec. The community began in December 2009 and, at the time of the study, had approximately 7000 members. The aim of the community is to provide technical support on a range of products and services. Thus, as with Super User (discussed in section 4.2), the majority of activity is goal driven – users seek answers to technical-support questions. The community is hosted by an external community software provider called Lithium[64]. Their service is typical of online Q&A communities but, although similar to Super User, lacks many of the sophisticated elements introduced with the Stack Exchange

[64] www.lithium.com

network. The forum software has a more generic design, for example, which could be used to service regular conversation as well as online Q&A. However, the general pragmatics of interaction remains the same. To ask a question, a user initiates a thread on a relevant board. To answer, a user submits a reply. If the reply answers the question to a sufficient degree, the original poster (or OP) will (or at least should) accept the reply as an answer. Other users who experience similar technical-support problems can draw on this answer, while also having access to the chain of reasoning that contributed to the answer. Users can reward "kudos" to other users in the community. Kudos, like reputation in Super User, is a way of rewarding a user who has asked an interesting question or submitted a well crafted answer, and helps to raise the profile of effective contributors while surfacing useful content in the system. The community is managed by a dedicated online community manger. Their role is to facilitate the community and to answer questions that are not addressed by other community members. Facilitation includes improving engagement, promoting effective contributors and converting passive to active users. While the aim of the community is to socialise the technical support process, Symantec employees are also active in the community. However, the preference is for community members, and not Symantec employees, to submit answers to technical-support questions.

5.2.2 Domain Analysis

As part of the introduction to the team, we presented some of our previous work. We described the aim of this work, which is detailed in chapter 4, and discussed, in broader terms, how we thought visualisation could support the role of the community manager[65]. In a way, much of this discussion was driven from the literature[66], as we had little working knowledge of how the community manager went about their daily routine. We had conducted some informal interviews with community managers from other organisations, but their approach to community management varied greatly. In some instances, their role was quite casual, and the process had not been formalised into strategic or operational directives. Also, in some cases, the notion of visualising a community to support community management had been met with little interest. They found it difficult to conceive of a use for such as tool. As a result, the aim of the domain analysis was for us to find out as much about the team's internal work practice as it was to elicit a set of design requirements for a visual analytics system.

During the domain analysis, we conducted three workshops in total. The first two workshops involved three members of the team, including the manager of the community and the author of this thesis, and were aimed at understanding the team's existing work practice and their need for visual analytics. The third and final workshop included an additional member of the team and a second researcher. This workshop enables us to present and get feedback on some initial designs and to identify a feature set to cluster users into representative cultures (discussed further in Sect. 5.2.4). While we wished to video each workshop, some of the team felt more comfortable with recording the session using audio. In this section, we present and discuss the findings from each of the three workshops.

[65] There is little in the academic literature as regards the process of formal online community management when conducted by a professional online community manager. There is a much stronger emphasis on the process of governance, or the way in which communities develop governing variables as opposed to the official or strategic process of online community management. There are some excellent practitioner guides however, notably Bacon's "The art of community"(Bacon, 2009) or O'Keffe's "Managing Online forums" (O'Keffe, 2008).

[66] While there is a growing body of work on the analysis of online communities, as illustrated by (Wagner et al., 2012), much of this research is very early stage or driven by questions that interest the academic community. How these sorts of measures, some of which are difficult to interpret, are adopted in real world scenarios is very much an open question. Furthermore, in many cases, evaluation strategies have not adapted sufficiently to include real-world scenarios or more grounded evaluation protocols.

5.2.3 Exploratory Workshop

The aim of the first workshop was to develop a general understanding of how the team underwent the practice of online community management and to establish how they thought visual analytics could support this practice. We divided the workshop into two stages. In the first stage, we addressed the following questions:

- What is the process of online community management?
- How do you currently analyse or evaluate community activity?
- How do you envision visualisation supporting the community manager?

Each question was considered open-ended and used to guide discussion. The author of this thesis facilitated and took notes. In the second stage, the team sketched out some initial design ideas on a white board, which we then brainstormed as a group. Here we describe both stages of the workshop.

5.2.3.1 What is the process of community management?

From the outset, it was highlighted that the process of community management was not to suppress or control the community, which as the literature highlights can reduce contribution rates (Preece, 2000a), but to support and encourage pro-social[67] behaviour. The aim, it was continually emphasised, is to reinforce, and not undermine, the sense of community. For example, the community manager mentioned that over the two years since the community started, they have only banned one user. Rarely has the community experienced anti-social behaviour to the point that it needed to be addressed on a broader scale or at a policy level. The team were more interested in increasing "conversion" rates. Conversion is the term used to describe the process of converting a user from "detractor" to "supporter". A detractor is someone who discredits a product or service or, possibly, contributes to the community in a negative way[68]. A "promoter", on the other hand, is someone who either promotes the company's products and services or contributes to the community in a positive way.

Aside from converting detractors to promoters, members of the team were also interested in improving "peer support" in the community. Peer support is the term used to describe when a user, who is not a member of Symantec's staff, answers a question belonging to another user in the community. Through these actions, peers provide technical support to the community, which, in turn, reduces the burden on the organisation's technical support department. Coupled with increasing conversion, improving peer support is considered as instrumental to the organisation's online community strategy.

Regularly, community managers will present cases to the management team about specific conversions (amongst other noteworthy incidents and events) in their respective communities. As described by Brown and Duguid, these presentations help to reinforce successful management practice by allowing the team discuss how best address difficult or troublesome situations (Brown & Duguid, 1991). Also, the idea of "conversion" is an identifiable practice that can be readily communicated to management, which helps to validate, and justify, their online community strategy.

---

[67] As identified by Butler et al., the term pro-social will be used to describe any behaviour that is considered socially beneficial, or contributes by way of value, to the community (Butler, Sproull, Kiesler, & Kraut, 2007).

[68] While negative contribution was mentioned briefly, the term detractor is primarily reserved for users who discredit Symantec products and services.

5.2.3.2 How do you currently analyse or evaluate community activity?

The community software provider that host Symantec's online communities provides a measure of community health called the "community health index". This is a weekly report that outlines the community's current standing. The team do not find this measure particularly useful. Rather they described the measure as too "coarse" or having "no value-add" and thus unrepresentative of how they understand the progression of their communities. It was also identified that community health is very difficult to measure, and that striving for a single measure, as opposed to a composite of different yet related measures is often meaningless. Nevertheless, currently, there is no generally accepted way for the community manger to assess status or evaluate activity. They did indicate that ratios, such as the thread-to-post ratio, can be used to make "snap judgements" on different users but that those judgements would need further validation. Users can have high thread-to-post ratio yet, qualitatively, contribute very little to the community. Conversely, low thread-to-post ratios may indicate "ranters" - users who are active on one thread but only post negative comments about a specific topic.

5.2.3.3 How do you envision visualisation supporting the community manager?

In answering this question, the team identified four high-level objectives for the system:

Objective 1: Detractors

The team were interested in "detractors". Detractors are users who post negative comments about Symantec products or services. Automatically identifying negative comments in social media communities is difficult however, due to the nuance of language and the context of use. While structural or social-network features, such as reply-to-networks, can help to identify cliques of dissatisfied users (through the analysis of detractor comments and voting patterns for example), realising the sentiment of the content in the first place requires some form of textual analysis on open corpus data. The difficulties of sentiment were discussed at some length. A member of the team, who had experience of sentiment analysis in previous projects, suggested avoiding this approach entirely due to the "inaccuracies" and difficulties that it presented in the previous project. They suggested that the information provided in the visualisations must be "reliable" and "actionable" and that, because sentiment is too nuanced a concept to address at this point, identifying detractors (if addressed at all) should be deferred to future iterations of the tool. Nevertheless, the ability to reveal colluders, users who support vocal detractors, was identified as important by a member of the team. Also the ability to reveal sock puppets, users who create alternate accounts with the aim of gaming the system or trolling other users, as discussed in (Gazan, 2009; Suh, Chi, Pendleton, & Kittur, 2007), was suggested as potentially useful by members of the team.

Objective 2: Peer Support

Part of the online community strategy is to increase peer support. Peer support involves members of the community answering questions about Symantec products and services. As discussed, this reduces the burden of effort on the community manager, or other officially appointed community members, which in turn reduces the cost of providing technical support. The team were interested in questions and metrics around peer-support, such as, how much support in the community can be considered peer support? And is change in peer support evident over time? They were also interested in identifying users that could contribute more answers to the community. One of the team mentioned that while they know their community very well, she felt that there could still be users with untapped potential that slip through the net as she was unable to identify their potential early enough. Being able to identify users that could provide a greater contribution to the community is important.

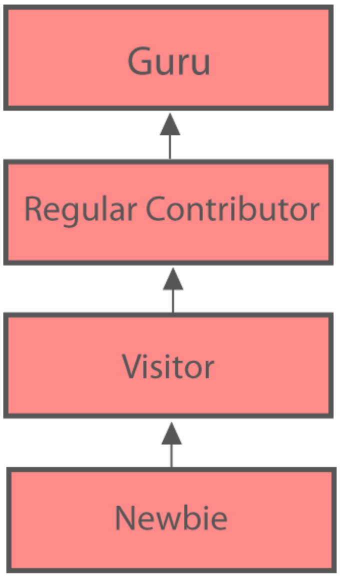


Figure 39: An illustration of community cultures, as presented by one of the team in the first workshop. In this image, a user starts out as a newbie. Over time, that user becomes a visitor, develops into a regular contributor and then a guru. This is an "ideal" trajectory, from the perspective of an online community manager, for a user in a technical support forum.

Objective 3: Cultures

The team was interested in partitioning the community into several different "cultures" with each culture representing a degree or level of contribution to the community (**DC3** discussed in Sect. 3.2.2). To take a simple example from the first workshop, as illustrated in Figure 39, the culture "newbie" could represent those that have just joined the community. The visitor culture could represent someone who visits the community regularly, asking questions but not contributing answers. The culture "regular contributor" could represent an active user who both asks and answers questions and the culture "guru" could represent the most knowledgeable users who are few in number but answer many questions. These sorts of demographics are common in Q&A forums, amongst other types of online communities, and have been discussed in Sect. 3.2.2. With this approach, the team could identify users that have the capacity to contribute to the community in a more meaningful way and attempt to convert those users from one culture to the next over a sustained period of time. Potentially, a user who starts out as a newbie could end up as a contributor or indeed a guru (to return to our previous example) through the encouragement and support of the community manager.

Objective 4: User Paths

The team were also interested in the notion of a user's path. A member of the team described a user's path as "a logical path, which is based on increased levels of activity". The team hypothesised the "Gurus" could have "ideal" paths and that these paths could be used to compare and contrast different user trajectories over time. If, for example, the community manager is interested in promoting a user to the position of Guru, they could use the path visualisation to confirm and thus validate their decision. Similarly, the path visualisation could be used to communicate this decision to the rest of the team or to upper management and thus illustrate the effectiveness of current strategy. Finally, comparing this approach to Digital Signal Processing (DSP), there was some discussion as regards the ability to automatically compare paths using techniques drawn from DSP. This could be particularly useful in large communities with several million contributors, enabling the community manager to identify and examine potential Gurus in a significantly large dataset. However, this was a preliminary discussion that helped to frame the approach.

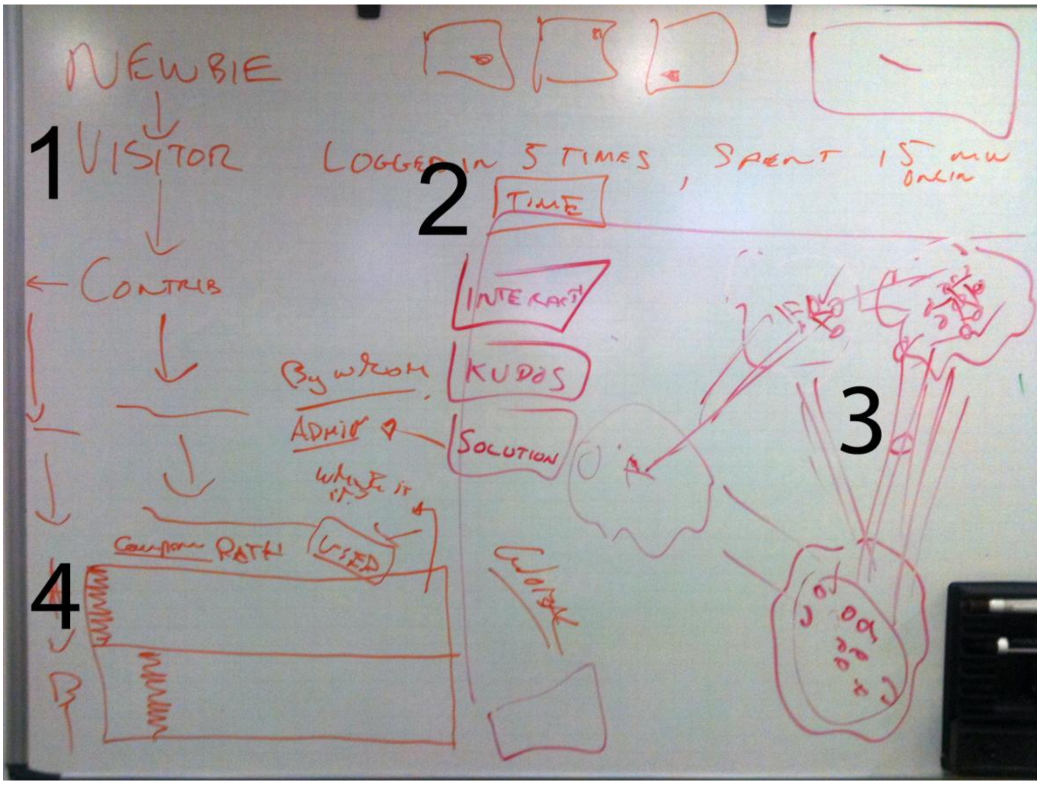


Figure 40: The results from the second stage of the first workshop. General ideas and initial designs as proposed during brainstorming. 1. Cultures as articulated by the participants in the first workshop. Cultures are defined by their levels of contribution, from newbie to visitor to contributor. 2. Analysing different user actions. 3. A team member's initial diagram of community cultures. 4. Initial sketch of user paths.

In the second stage of the workshop, the team sketched some ideas on a whiteboard that were then discussed as a group (see Figure 40). We concentrated on visualising community cultures, which would provide a high-level "overview" of the community, and then user paths, which would describe a user's trajectory through the community. It was decided that the community culture visualisation would be the community manager's first contact with the visualisation system. From there the community manager could select users that are of interest and then "carry their analysis forward" by comparing user paths using other methods. Coordinating these visualisations **(DC21)** was suggested but discarded, due to screen real estate, the potential complexity of the interface and the quantity of information presented to the user.

5.2.3.4 Implications for the practice of online community management

Certainly, the outcome of these workshops could have implications for the role of the community manager and the practice of online community management, particularly in terms of visual analysis and reflective community design. Firstly, they will be able to extend the practice of "conversion" to other community cultures beyond that of detractor. Secondly, they will be able to assess, in an approximate fashion, the outcome of any interaction with the community. For example, if a certain activity is targeted at a specific set of users, and that activity is promoted and rewarded, what is the rate of conversion of users to the next culture? This will allow them to answer questions such as, who should be promoted? How successful was that interaction with the community? Are there gold standard approaches to encouraging certain users or dealing with specific cultures? Finally, visual analytics will be considered a key component in community management process, enabling the community manager to not only analyse the activity of the community but also assess their own actions, and the result of those actions, on the community. This approach opens a dialogue between the community manager and the community manager's practice by enabling the community manager to reflect on their actions in a continuous manner. In essence, visual analytics applied in this context encourages a more "reflective practice".

5.2.4 User Requirements and Design Considerations

Having completed the first workshop, we drafted a set of user requirements. In this section, we discuss these requirements in relation to the design considerations presented in Table 8. Where appropriate, we match each user requirement with a design consideration.

R1 Multiple Linked Visualisations – The team suggested that Petri consist of both global and local visualisations. The global visualisation provides access to the entire community, while the local visualisation allows the user to drill-down into data and compare different user paths. In this mode of analysis, the user shifts between explorative and confirmative tasks. First they explore the global visualisation to select users of interest and then confirm their selection using the local visualisation. Over time, the global visualisation can support both kinds of tasks, confirming that a user has migrated into a new culture while also allowing the community manager to explore the community. In this way, the manager can build a more comprehensive mental map of how the community is organised and how specific users contribute to the community. As mentioned in Sect. 5.2.3.3, we considered integrating the visualisations with a coordinated user interface **(DC21)**, but the team thought that this approach appeared complex and that some visualisations required the entire display. They preferred the idea of linking different visualisations into a single visual analytics system **(DC22)**.

R2 Cultures – The global view should automatically cluster the community into meaningful representations called "cultures". Each culture should represent a collective position in the community. Although we considered clustering on structural position **(DC2)**, partitioning users into a series of clusters based on their level contribution coupled with the level of peer recognition was required **(DC3)**. This is a linear trajectory that extends from users who are barley active in the community to the highest and most well respected contributors.

R3 Paths – The team were interested in visualising a user's path, their journey or trajectory, through the community. This is "a logical path, which is based on increased levels of activity". Given the need to compare and contrast paths, as a result, we proposed the use of small multiples **(DC25)**.

R4 Networks – The team indicated an interest in networks. Their interest was focused on the notion of a detractor - users who denounce or complain about a product or service. They wanted to understand whether detractors were supported by other users in the community. If a detractor can be identified, their complaints can be addressed and they can be encouraged to participate in the community in a more productive way. Cliques of detractors could present further difficulties for the community manager also. In considering the approach to representation, using matrices or node-link diagrams, the team expressed more familiarity with node-link diagrams but also recognised the difficulty in representing large networks. As a result, and in consulting in Table 8, we decided to focus on the ego and provide an approach that is simple **(DC27)**, clear and visually perceptive **(DC12)**.

R5 Outliers – As often required of visual analytic applications, community managers are interested in identifying, and if necessary filtering for, outliers. In this case, an outlier is someone on the periphery of a particular culture, or someone who has just moved from one culture to another.

R6 Interaction – The ability to interact with the application, to adjust the scope of time and filter for specific users, was also highlighted. Animation was discussed at some length and while considered superfluous to the analytic process[69] was suggested as an engaging way to interact with the system **(DC18)**. Parameter fine-tuning, whereby a user adjusts the feature set or the parameters of the clustering algorithm, was discussed but considered as excessively complex and would require further investigation beyond the scope of the current study. However, interactive exploration **(DC19)**, using pan and zoom was considered as a way to traverse the cultures visualisation, while search and filter enabled the identification of outliers or users of interest (3.5).

R7 Profile Information – Aside from providing access to the user's pattern of activity, the community management team also indicated a need to access the user's general profile information such as number of posts, time of registration, time of last visit etc. For labelling community users, we considered the use of social media icons **(DC11)**, given their familiarity to the community management team.

R8 Selection – Given the focus on linked visualisations, the team requested the ability to select users, which could then be "carried forward" for further analysis.

R9 Change – The team were also interested in identifying changing patterns in their community. They were primarily interested in identifying population shifts between cultures. Given that in the previous study the use of animation was not particularly effective, we proposed the use of small multiples **(DC25)**. However, we also suggested that this requirement could be postponed to later iterations of the tool.

R10 Design– Finally, the team suggested that the application should have a certain design aesthetic (Sect. 3.3). This was considered important if the tool was to be used as part of a broader online community strategy. Similarly, and from a more practical perspective, they asked that the application be web-based or have the ability to run on a tablet. This is because managers use tablets to quickly share observations and insights with other members of their team. Thus, a web-based approach would help improve the potential for adoption and the dissemination of findings. While these are practical considerations, and usually reserved as implementation details, their inclusion illustrates the requirements needed for the adoption of visual analytics into regular work practice by novice or more casual users.

5.2.5 Second and Third Workshops

In the next two workshops, we focused on producing an initial set of designs based on the design requirements and selected design considerations, validating those designs as mock-ups presented to the team, selecting a set of features for the clustering process and then validating the clusters with the community manager.

5.2.5.1 Second Workshop

In the second workshop, the author gave a brief presentation on relevant approaches to information visualisation. All of the examples that were used in the presentation were based on the requirements gathered during the first workshop. Not all examples were specific to online communities; however, several of the visualisations drawn from chapter 3 helped to illustrate the gamut of approaches currently used to address community datasets. Having completed the presentation, the author handed out an information sheet (Appendix C) that detailed different ways to visualise a community's culture or a user's path. The sheet presented an illustration of the visualisation on the left

[69] Certainly it appears that the visualisation community remains uncertain as to whether animation can enhance the analytic process (Archambault et al., 2010; Bederson, 1999; Robertson et al., 2008; Tversky et al., 2002). This is considered in more depth in Sect. 3.5.

and a textual description of the visualisation on the right as illustrated in Figure 41. The information sheet is available in Appendix C.

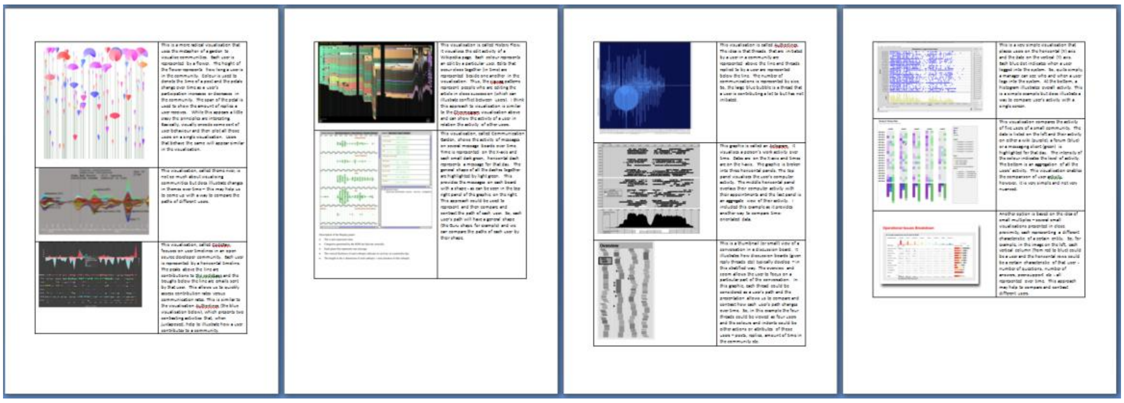


Figure 41: A sheet of visualisation examples used to guide discussion and present alternative approaches during the second workshop (Appendix C).

The team addressed each visualisation in turn, with each member providing their opinion of the approach, and then we discussed the advantages and disadvantages of each visualisation as a group. The aim was to see if there was a preference for a certain type of visualisation outside of what had been discussed previously. Also, it provided some alternative representations that the team had not previously encountered but could help understand the problem from a different perspective. The rest of the workshop was used to address some initial sketches, to clarify the different interactive techniques and to discuss possible alternatives.

5.2.5.2 Third Workshop

The third workshop had two specific goals: To gather feedback on our initial designs (as illustrated in Figure 42) and to identify a set of features that could be used to cluster the community into meaningful cultures. We presented the mock-ups to the team, which involved providing the rationale for our design decisions, and held a discussion on the approach to clustering. Some initial work on the clustering had been completed at this point, so the general idea was presented and discussed at some length. Communicating the idea behind the clustering was also discussed. Obscure measure renders the representation less comprehensible it was suggested, thus it was important to create a model that community managers could understand. Following the workshop, we drafted up all the metrics, numbering over 60 in total, which could be used as features for clustering. This was then reduced to a smaller set as discussed in Sect. 5.3.1.1.

As regards the visualisations, the team felt comfortable with how the cultures were represented (see Figure 42). At the time, there was no discernible order to how each culture was positioned on the screen. The team suggested that the position of each culture should convey some meaning or have some information value. One member of the team suggested that the ordering of the cultures should be represented in the visualisation – from newbie to guru for example. They had mixed views on representing lurkers (which are the largest of the cultures in Figure 42). One member of the team thought lurkers were the most interesting culture but that they occupied a lot of screen real estate and commanded a lot perceptual attention if they are not to be used in the actual application. The paths visualisation was less well received. Initially, user activity was broken into four actions, kudos received, kudos given, questions (initiated threads) and answers (replies). The user's path was divided horizontally into time units – days, weeks or months – and each time unit was divided vertically into actions. The number of actions was illustrated using colour intensity. The advantage of this approach is that, in one timeline, the reader can compare the actions of different users; however the community managers found the composition as non-intuitive and overly complex and it was suggested that a more familiar approach be considered. As result, we discussed timelines, line graphs and small multiples as an alternative and more familiar representation.

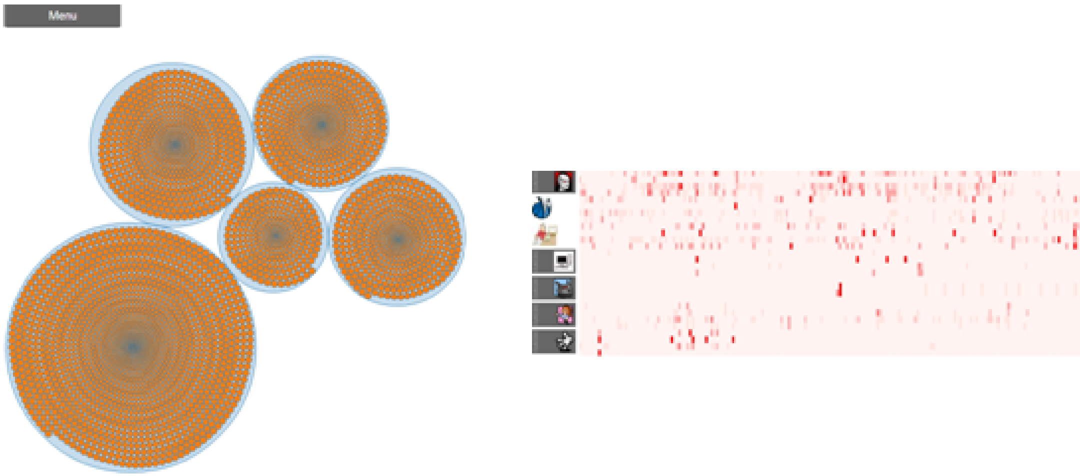


Figure 42: Some initial mock-ups. The cultures visualisation (on the left) is close to how the cultures are represented in later versions of Petri. The paths visualisation (on the right) was rejected as noisy and non-intuitive. In this visualisation, the users are represented by profile images on the left and their activity is presented like a heat map stretching left to right.

5.3 Implementation of Design Considerations

In this section we describe each visualisation, justifying our implementation with reference to the user requirements and design considerations discussed in the previous section. We also present our approach to partitioning and ordering the community into number of discrete cultures. Petri includes three different visual representations linked (DC22) into a single visual analytics system with each visualisation takes up the entire screen display. A user of Petri is able to select users and then consider their activity from several perspectives - part of a culture, compare their paths or analyse their ego network.

5.3.1 Visualisation 1: Cultures

In this section we describe the implementation of the cultures visualisation. We start by describing the analytic abstraction, using a combination of clustering techniques (Sect. 5.3.1.1). Next, we discuss our approach to visual mapping (Sect. 5.3.1.2), or encoding the data as visual attributes, describe the use of interactive exploration to support pan and zoom (Sect. 5.3.1.3) and linked visualisation as a way to identify and analyse community users from different perspectives (Sect. 5.3.1.4).

5.3.1.1 Analytic Abstraction

Initially, we followed the approach proposed by Chan and Hayes, as discussed in Sect. 3.2.2, in which they use a number of social network measures, coupled with activity ratios, to cluster and partition a large online community into a number of different user roles. However, with this approach there is no indication of a clear trajectory, which users can follow from lurker to guru because, as identified in **DC2**, the categorisation is skewed by the structural position[70] of the user in the community. Consequently, we removed the social network measures, and concentrated,

[70] As considered in **DC2**, using social network measures take into account the users structural position in a community. Who are they closest to, who do they interact with in the community, are questions that these measures can help to answer. However, this was not applicable to providing a categorisation schema based on contribution rates **(DC3)**.

instead, on identifying a number of activity based ratios that could be used to illustrate an increasing level of contribution. We focused on ratios that indicate productive contribution, such as number of replies submitted to a thread that the user did not start, and peer recognition, such as the amount of kudos attributed to a user over-time.

The company's software provider collects over 60 metrics on each member of the community. These metrics range from the conventional, such as number of logins and number of minutes spent online, to the more obscure, such as arbitrary points and view productivity. As with most forum software, certain metrics are immediately useful, such as number of posts or number of replies versus number of initiated threads, however, we validated our feature selection with participants from the community management team (following the third workshop Sect. 5.2.5.2). With the help of one of the participants, we reduced the 60 metrics to 15, and identified 7 of those 15 as supplementary, and to be used only if we had difficulty in producing meaningful clusters with the initial set. Adding more features to clustering will not, however, guarantee more meaningful results, but that the supplementary features could be used if necessary to replace some of the initial set.

The resulting 15 features were:

Table 12: The final set of primary and supplementary features drawn up with a member of Symantec team.

	Primary Features	Explanation
1.	Initiated Threads	Total number of threads initiated
2.	Posts	Number of posts
3.	Message views	Total number of message views
4.	Replies	Number of replies
5.	Accepted solutions	Number of accepted solutions
6.	Minutes spent online	Total number of minutes spent online
7.	Registration time	Time of registration
8.	Kudos received	Kudos received by other users in the community
	Supplementary Features	Explanation
9.	Logins	Total number of logins
10.	Solved threads to threads ratio	Threads to solved threads ratio
11.	Total posts per thread	Number of posts per thread
12.	Solutions marked	Number of posts that are marked as a solution
13.	Number of private messages sent[71]	Number of private messages sent
14.	Last visit time	Last time the user visited the site
14.	Kudos given	Kudos attributed to other users in the community

As with the majority of online community populations, the distribution of activity is heavily skewed. Generally, this resembles a power law (called Pareto's law) in which roughly 80% of all posts in a community are contributed by 20% of the population (Anderson, 2006; Morzy, 2009). Some communities are more heavily skewed than others and distributions can range from 90%/10% to 99%/1%. Essentially, in accordance to DC3, our aim was to remove

[71] It is important to note that this is only a flat metric. We did not address the content of any private message or the reply-to-graph of any private message network. The team were **only** interested in whether the user sent private messages or not. A high value could signal a degree of commitment to the community, as the user is engaging in private discussions with other users, while a low number indicates someone who is less engaged. Of course, in either case it is more signal of intent, or social signal as discussed in (Donath, 2007), than a reliable or definitive metric.

marginal users from the population, thus “cutting the tail” of the long tail distribution, and then partitioning the remaining distribution into a number of discrete yet ordered “bands”, with each band representing an increasing level of contribution. This reduces the noise for clustering, but also helps the community manager to focus on the users who are active in the community. First, we choose users who submitted more than three posts to the community, which reduced the population to 879. Next, we experimented with a combination of the primary features in Table 12. We removed registration time, which was suggested as a way to incorporate a temporal dimension into the clustering process and was not required for this iteration, and converted the other features into percentages of the overall population. So, for example, the percentage replies of a specific user was calculated by comparing the number of replies that the user submitted to the community against all replies submitted by all users to the community, while percentage of kudos received is calculated against all the kudos that is distributed in the system. This allows us to situate each user in relation to other users in the community. We created one additional feature - the percentage of threads that a user contributes a reply to, which they did not initiate (shown in Table 13). This is an important addition as it indicates peer-based activity - users contributing to threads that they did not initiate with a question. We choose principal component analysis to reduce the dimensionality of the dataset and then used agglomerative hierarchical clustering as it does not require pre-defining the number of clusters. A similar approach was first applied by Chan and Hayes to decompose communities into user roles, as discussed above, and is detailed in a practical way in (Husson, Le, & Pages, 2010).

Table 13: Cluster statistics. The middle columns provide the means of the nine features for the 19 clusters. The first column is the cluster number and the last column is the population of the cluster. N.B. Cl in the first column stands for the “Cluster”, i.e. the cluster number.

Cl	Replies	Accepted Solutions	Initiated Threads	Kudos Received	Posts	Views	Minutes	threads	Non-initiate	Size
1	0.75	0	0.25	0	0.016878	0.0006482	0.0086079	0.02281022	0	2
2	0.75	0	0.25	0	0.0168776	0.00055733	0.0084206	0.03910323	0.0162930	7
3	0.7365501	0.01617632	0.2755102	0.00070837	0.0168776	0.0006183	0.0126181	0.033982571	0.00844779	49
4	0.7398325	0.01544103	0.2601674	0.0010518	0.02132720	0.00109018	0.01671008	0.0468646	0.0182481	55
5	0.7423309	0.01993561	0.2576691	0.00260733	0.0281690	0.0013686	0.02034079	0.05718662	0.02377403	71
6	0.77012406	0.00941295	0.2298759	0.0013963	0.02068134	0.00091810	0.0143510	0.0374177	0.0126973	203
7	0.78590876	0.010722941	0.21409125	0.00214747	0.03091037	0.00192022	0.0395176	0.08266423	0.03576642	132
8	0.7936631	0.0249115	0.20633694	0.00337846	0.04587342	0.00241718	0.0395176	0.08266423	0.0357664	125
9	0.7985033	0.03023245	0.2014966	0.00370691	0.0555487	0.0035999	0.0385052	0.1067430	0.04384877	103
10	0.8492664	0.0557137	0.1507336	0.0099084	0.0794506	0.0006438	0.0767692	0.1494797	0.088086	94
11	0.873083	0.2251818	0.1269164	0.0491727	0.349252	0.0205566	0.2371149	0.9103351	0.6822329	22
12	0.9675596	1.6749233	0.0324404	0.293108	0.7264416	0.0454212	0.9015536	14.94069	2.4026764	6
13	0.986842	0.070771	0.013158	0.02314	0.962025	0.051323	2.087226	1.117701	1.04927	1
14	0.959016	1.98159	0.040984	0.485942	1.544304	0.059372	1.577933	4.425182	4.083029	1
15	0.833368	2.642132	0.166323	0.539935	2.41490	0.084248	0.964431	7.010341	4.77794	3
16	0.991286	5.095541	0.008143	0.766516	3.911392	0.221030	6.677768	13.20712	13.01323	2
17	0.816449	5.732484	0.183551	2.296656	4.206751	0.583894	8.340355	14.59854	10.42427	1
18	0.987507	15.28662	0.012493	1.700798	8.105485	0.315316	3.804482	24.74909	24.20164	1
19	0.989204	18.25902	0.010796	2.302441	9.379747	0.52056	8.977342	22.71898	22.17153	1

Clustering partitioned the community into 19 different clusters as described in Table 13. Next, we evaluated and then aggregated the clusters by the commonality of their features resulting in a set of easily describable cultures. These clusters had to be communicated to the user of Petri so the aim was not produce too fine-grained a representation. Although we used agglomerative hierarchical clustering, which can be used to organise the clusters hierarchically, we choose to conduct this task manually as the initial set of clusters provided an accurate representation.

The process resulted in 4 cultures:

Pragmatist (Clusters 1 - 7): Users that ask a question and then receive an answer. Generally, they do not contribute to other threads that they have not initiated. They receive very little kudos and are mainly in the community to seek an answer to a question. These users make up the majority of active users in the community.

Peer (Clusters 8 - 10): Users that contribute to threads that they have not initiated. However, they receive little recognition from their peers, by way of kudos, and have lower contribution rates than the apprentice culture. This group shows some potential.

Apprentice (Clusters 11 - 12): Users that make a recognisable contribution to the community, participating in more threads than they initiate, thereby contributing more answers than questions. They have also begun to receive recognition in the form of kudos, views and accepted solutions from the rest of the community.

Guru (Clusters 13 – 19): Finally**,** gurus contribute the most to the community. They are highly knowledgeable and skilled users that receive the most peer recognition in the form of the kudos, views and accepted solutions.

5.3.1.2 Visual Mapping

Initially we choose a scatter plot, varying the hue of each visual variable according to their culture. However, the resulting visualisation was sparse and required the use of “jitter” to reduce occlusion. Further, the visualisation posed some interesting questions as to how the tool would be used in practice. It was felt that a scatter plot would be more suited to the sorts of explorative analysis performed by a data analyst, or as described by Bertin at the point of “inventory and classification” (Bertin, 1983a), as opposed to a tool that is used at regular intervals by a community manager. In the latter case, the dataset remains the same and the task may shift from explorative to confirmative as time progresses (in addressing simple questions, such as, have we more gurus this week than last? Are there more pragmatists this week than last week? (These are two community cultures)). Even having enclosed each culture using convex hull (and quadratic spline for smoothing), the occluded and dispersed data points remained problematic. The community manager would have to seek out each culture when addressing the tool and comparing cultures, as regards their size and population, would also be difficult. The approach, we observed, would not help the user build a sufficient mental map of the community, particularly as the tool is to be used at regular intervals over time.

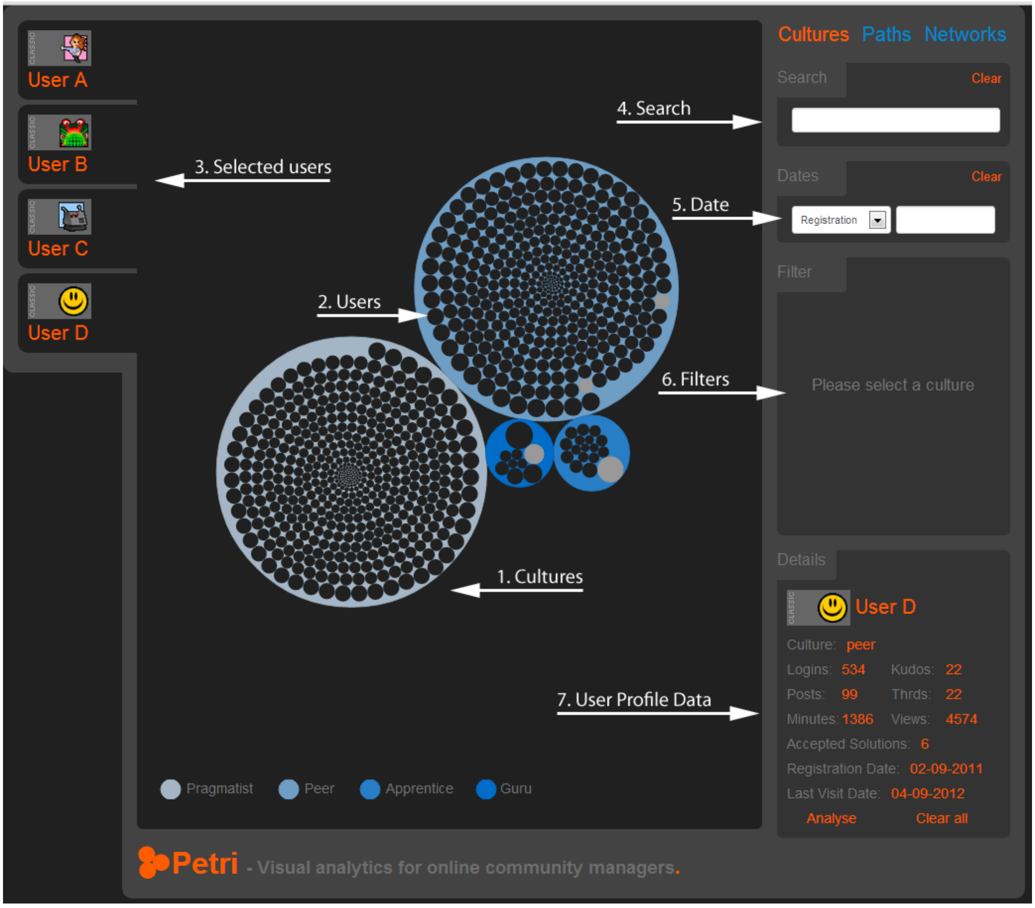


Figure 43: Petri Cultures Visualisation. 1. Interactive dispaly with different cultures organised in a clockwork configuration: pragmatist, peer, apprentice and guru. 2. Individual users, organised according to the distance from the medoid of their cluster. The larger the glyph, the further the user is from the centre. Exceptionally large glyphs are outliers. 3. Selected users. These users are "carried forward" by the communtiy manager for analysis. 4. Search functionality. 5. Date selector. 6. Filters (only available when the commuty manager zooms into a culture). 7. User Profile Data (R7).

Comparing change **(R9)** in cultures over time is not addressed iteration of the tool but, as discussed, this requirement contributed to the choice of technique. The cultures visualisation, given the static position of each culture and the reliance on size to illustrate population, can be used in a small multiple display to illustrate change in the community's structure over time **(DC25)**. This is not reported in this chapter however; as we did not address it in the evaluation.

We re-organised the layout and composition of the visualisation. We drew together each culture, as illustrated in Figure 43 and ordered each culture in a spiral (clockwise) orientation according to their level of contribution (From Pragmatist to Guru). Thus, each culture has a fixed spatial position. We varied the saturation of each culture (using a degree of blue) and provided a legend. We used a circle packing algorithm, as proposed in (Wang, Wang, Dai, & Wang, 2006) to avoid occlusion, thus ensuring that each user is perceptible. This approach also accounted for interaction and guaranteed that the community manager could select, and, thus, retrieve the details of a user at any point **(R8)**. Also the community manager could filter different users based on last visit or registration date or search for a particular user with a given username. We sized and ordered each user according to their distance from the medoid (centre of their cluster), thus outliers appear larger and to the outskirts of the culture **(R5)**, than closely related users, who are packed together at the centre of the culture. Although the size of a visual variable has less efficacy than its spatial position, which is generally considered the "fundamental substrate" of visualisation (Munzner, 2000), we felt this approach was appropriate, given the requirements of the management team. One drawback, however, is that the community manager is provided with no information on how closely a user may be related to a second culture or how a user is related to a user from a second culture.

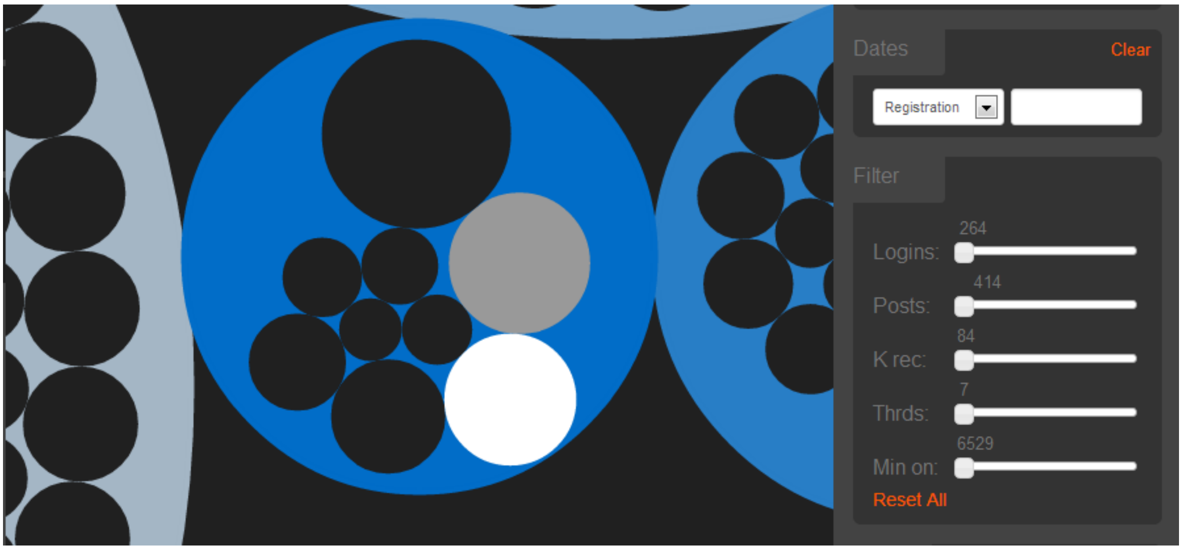


Figure 44: Cultures Visualisation. Zoomed in on Guru Culture. Users filtered on login.

5.3.1.3 Interaction

We sought to provide engaging yet intuitive ways for the user to improve the readability and representation of the visualisation **(R6)**. Thus, the user can pan and zoom (using interactive exploration **DC19**), search (using auto-suggest) and filer out users on a range of different attributes (number of logins, posts, kudos, threads, views and date or registration or date of last visit). Filters can only be applied once the community manager has zoomed into a specific culture (see Figure 44). This is because there is a significant gulf between the contribution rates of users in the different cultures. For example, one of the top contributors, a guru, logged in over 3050 times at the time of writing. This is in contrast to the majority of pragmatists who have logged in less than 30 times. Filters can be combined to isolate users within a specific set of parameters, for example, users that have received a certain level of kudos combined with a certain number of posts. The overall approach reflects Shneiderman's highly cited visual information-seeking mantra, overview first, pan, zoom and filter, details on demand (Shneiderman, 1996) and allows the manager to traverse and explore the community in a intuitive manner. While we discussed the possibilities of parameter adjustment, reordering and automatic clustering in the workshops, we decided against it in this iteration. The results of these techniques can be difficult to interpret and can require much trial-and-error on behalf of the user, which is something we aimed to avoid. Further research is required before implementation and evaluation of this level of interaction.

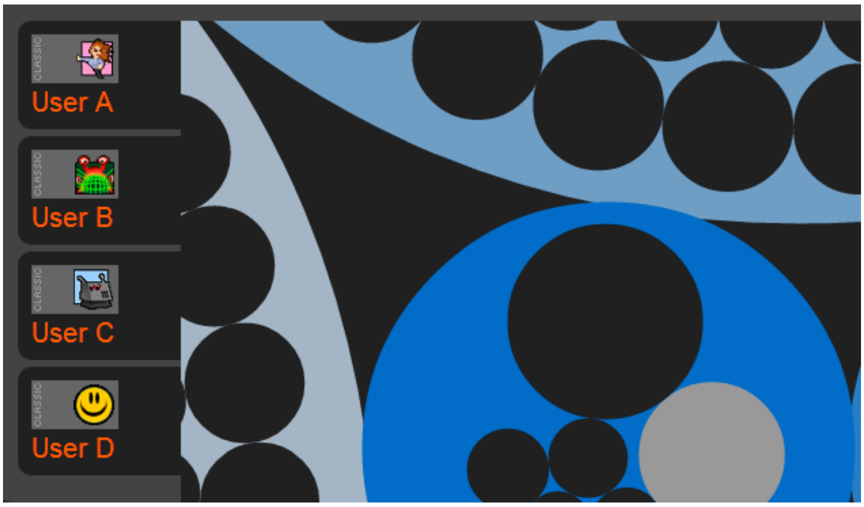


Figure 45: Selection (R8) - Users selected for analysis by the community manager.

5.3.1.4 Linked Visualisation

In Petri, we implemented a relatively simple selection function to fulfil the requirement **(R8)** and provide the community manager with a way to select users for further analysis. This results in a linked visualisation discussed **(DC22)**. Selected users are presented to the community manager in the top left of the visualisation interface (see Figure 45) and are persisted across the cultures, paths and network visualisations. At both the cultures and networks interface, the community manager is able to select additional users for analysis.

5.3.2 Visualisation 2: Paths

In this section we consider the implementation of the paths visualisation. In particular, we discuss the approach to visual mapping (Sect. 5.3.2.1) and how we used interactive animation to maintain orientation when traversing between a line graph and stacked graph (Sect. 5.3.2.2).

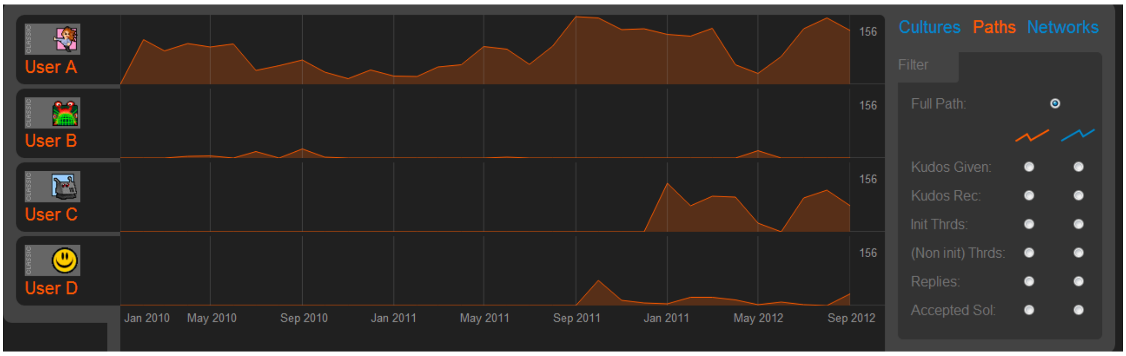


Figure 46: Path visualisation. In this image, the communtiy manager is comparing the combined contirbution rates of four users. The user's path (considered as an aggregation of all productive activtiy in the community) is presented on a time-series. At a glance the most active contributors are perceptable, however, the approach also allows the community manager compare their contribution rates with other users.

5.3.2.1 Visual Mapping

The paths visualisation, as discussed in the Sect. 5.2.4, enables the community manager to analyse and compare how users contribute to the community over time **(R3)**. A user's path is an aggregation of any recorded productive action that is committed during their time in the community. Other actions, such as giving and receiving kudos, are also included in the user's path as these actions signify social exchange and peer recognition by the community as a whole. So, for example, initiating a thread or replying to a thread, are both social actions that contribute to a user's path, as are having a solution accepted, giving kudos, receiving kudos and posting to threads that the user did not initiate (as this is considered peer-based behaviour). Initially, we considered the user path as an invariant represented by two components - aggregated actions over time (Bertin, 1983b). We used a time-series to represent this information, reducing the number of categories in the time component to intervals of four months (since the community began). We avoided cumulative frequencies so that the graph would present the general tendency of a user's contribution since joining the community.

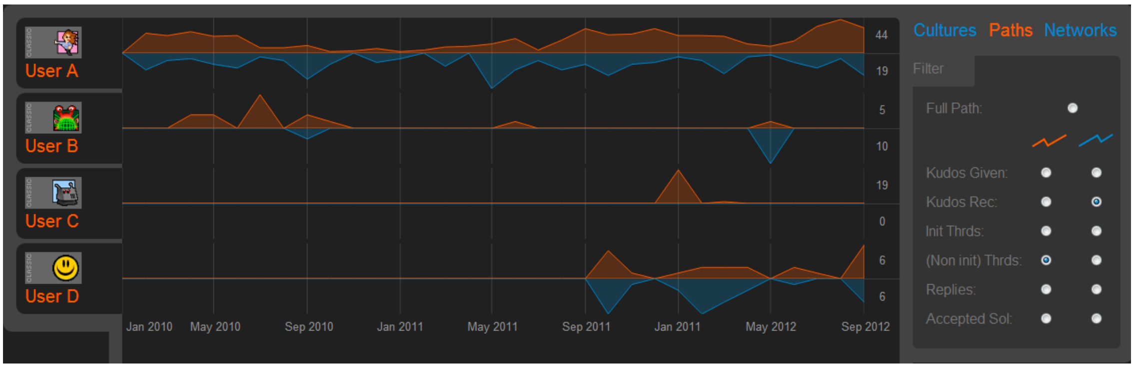


Figure 47: Filtered path visualisation. In this image, the community manager is comparing the paths of four different users. Non-initialised threads are in orange and received kudos is in blue. The representation enables the community manager to quickly evaluate how useful a user's contribution is considered by the community over time.

5.3.2.2 Interaction

While this approach facilitates comparison, it is quite limited in scope. The community manger is unable to assess how exactly an individual user contributes to the community. They are unable to answer questions such as: Does this user post replies or initiate threads? Or, do their replies receive kudos from other users? Further, a user's path may be heavily skewed by a single action and that single action is not presented to the community manager in the context of the user's other actions. However, to provide this context, we had to re-address the construction of the visualisation. Our objective was to provide a second visualisation that would draw from the first, and thus have a similar design aesthetic, yet provide the ability to compare and contrast a user's actions over time. For example, a community manager could asses a user's replies versus their kudos received or their number of posts versus their number of replies. To do this, we had to add a third component to the visualisation in a way that would not (excessively) increase the complexity of the visualisation and thus reduce retention. We drew some inspiration from CodeSaw, as discussed in Sect. 3.7.1, and introduced a third component (see Figure 47) as an inverted line graph positioned under the x-axis. Now the visualisation resembles a stacked graph (Jeffrey Heer et al., 2010). Using the interactive filters **(R6)**, the community manager is able to "drill into" a user's path, reducing the path to their individual actions and then comparing those actions over time. Interactive animation **(DC18)** is used to transition between the line graph (Complete Paths illustrated in Figure 46) and the stacked graph (Individual Actions illustrated in Figure 47) and hence maintain orientation.

5.3.3 Visualsiation 3: Ego-centric Networks

In this section we discuss the implementation of the network visualisation in terms of visual mapping (Sect. 5.3.3.1) and interaction (Sect. 5.3.3.2).

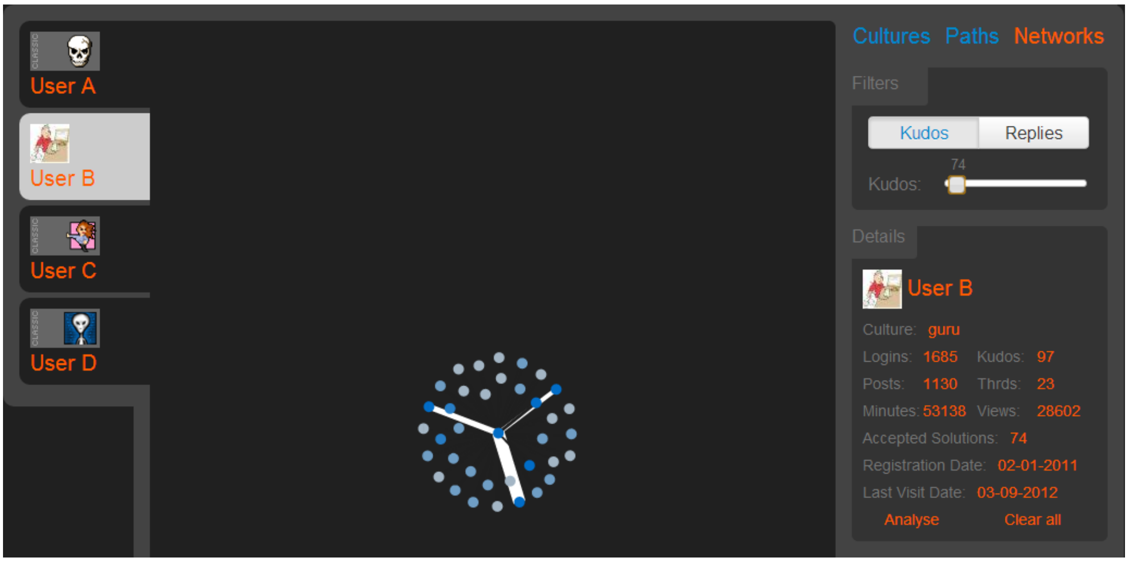


Figure 48: Network Visualisation. In this image the user is analysing the ego-centric networks of four users. They are looking at their kudos network and has filtered out any connections that are fewer than 74 kudos.

5.3.3.1 Visual Mapping

Community managers are also interested in networks **(R4)**. The use of networks as a way to visualise the community was proposed by a member of the team in the first workshop. However, the issue of addressing large graphs was also raised and has been considered in chapter 3. As opposed to visualising the entire graph, Petri provides access to the ego-centric (directed and 1 degree) network of selected users (Fisher, 2005). This approach enables the community manager to explore the immediate social network of any selected user, from the dual perspective of the kudos and the reply-to graphs. This helps the community manager to answer questions such as "Who kudoed this user?" or "Is there significant communication between a subset of users?" or "Have two users unusually high levels of kudos exchange?" Applying relatively simple network analysis in this way furnishes the community manager with another perspective from which to consider the interactions of their community but also ensures that the graph remains legible **(DC9)**.

5.3.3.2 Interaction

Petri allows the community manager to filter out users based on the number of connections **(R6)**. So, for example, the community manager can focus on a user who has been "kudoed" by another user over a particular threshold (see Figure 48). This enables the community manager to isolate users based on specific communication patterns. While this approach does not attempt to automatically identify detractors or sock puppets, as was discussed in the first workshop, it does allow the community manager to confirm their suspicions about particular patterns of user activity. This can be a powerful tool in combating against negative or potentially disruptive behaviour in the community. The community manager can also select other users from the network and examine then their paths and networks **(R8)**.

5.3.4 Implementation Details

Petri was implemented using the D3 JavaScript visualisation library[72]. The Ruby on Rails web framework, or more specifically the rails-api, was used to provide server support and a MySQL database[73] was used for persistence. Rails-

[72] d3js.org

[73] dev.mysql.com

api[74] is a middleware subset of ruby on rails that provides data (primarily JSON) API access without the addition of other functionality such as html and JavaScript templates. Symantec's community forums are hosted by Lithium that provide data access via a Restful API. Having implemented and evaluated the approach to clustering, we made daily syncs between the local and remote databases using this API. As a result, all trials were conducted on a (almost) live dataset.

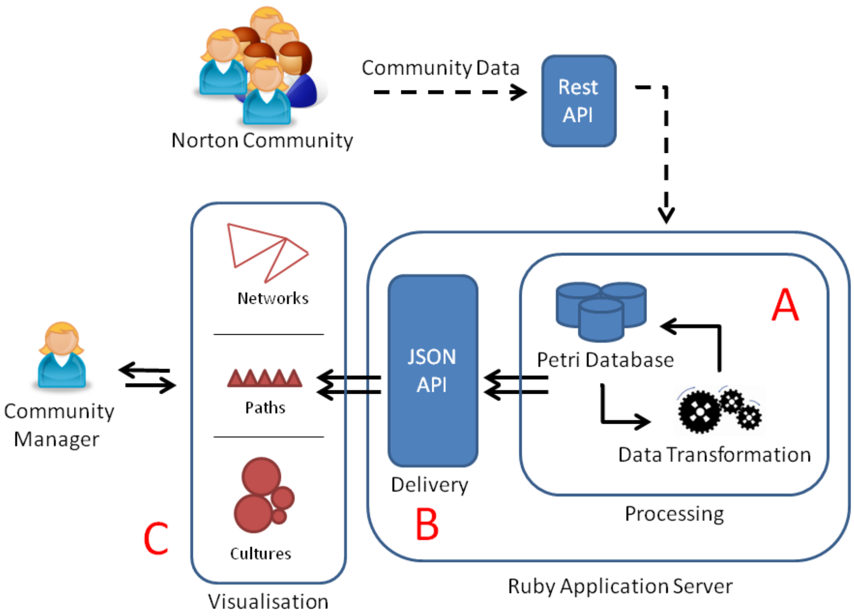


Figure 49: *Petri* Component Architecture

There are three major components. The first is a data Processing component (see A on Figure 1) that takes data from the community REST API, enters that data into a MySQL database and then processes that data using the clustering algorithm to generate several clusters based on the user activity / contribution to the community. The second is a Delivery component (see B on Figure 1) which is JSON REST API written in Ruby on Rails. The API provides data access to the community dataset which has been processed in the data processing component. The final component is the Visualisation component (see C on Figure 1). The Visualisation component is a JavaScript application built on D3 and jQuery which takes data from the JSON API and visualises it for the user. The visualisation component has three different visualisations:

- Visualisation 1 (section 5.3.1) is a set of clustered users organised according to their contribution to the community. The visualisation is organised using a circle packing layout to minimise on space and ensure a clean representation for the community manager.
- Visualisation 2 (section 5.3.2) enables the user to compare different user paths in the community. A user path is considered as a user's absolute contribution to the community over time.
- Visualisation 3 (section 5.3.3) is a network view that shows a user's ego-centric network.

[74] github.com/rails-api/rails-api

5.4 Evaluation of Design Considerations

As with the previous case study, by selecting a set of design considerations from Table 8 and then implementing these decision considerations as elements in a system we are able to evaluate their applicability within a single community context. This approach allows us to establish their feasibility and to consider the implementation with respect to Table 8. However, as argued throughout this thesis, it is important to evaluate the system in a grounded strategy with representative users. From this perspective, we evaluated Petri with five members of Symantec's community management team (2 women and 3 men). The evaluation took place over a single day at the Symantec offices in Dublin. Three of the team, who participated in the evaluation, were involved in the initial participative design workshops. The other two were aware of the project but had not participated previously. All five participants were familiar with the community; however, one of the participants was more involved with Symantec's enterprise communities than the company's technical support forums.

5.4.1 Procedure

The five sessions were conducted back-to-back and each session was videoed for analysis later. First, we asked some preliminary questions to establish the participant's familiarity with the community and their general understanding of visualisation and visual analytics systems. Next we asked them to complete some simple explorative tasks. In previous studies we observed that, even with highly motivated participants, loosely described yet open-ended tasks help guide the participant through the evaluation. Asking the participant to simply "explore the visualisation and tell me what you observe" can, at times, leave participants uncertain about what to do next which in turn require intervention from a facilitator, which is particularly the case with interactive visualisation. However, very narrow tasks can have the opposite effect, as the participant does not explore or learn from the visualisation, and, as a consequence, the results are limited to the prescribed tasks (North, 2006). We defined five loosely-focused and open-ended tasks to help guide each participant through the evaluation (Appendix D). They were informed that each task was considered explorative, there was no correct answer, and that we were not observing completion time. Also, if they wished to pursue their own interests or objectives, irrespective of the tasks, we suggested they do so. Each participant was encouraged to think aloud and describe any **insights or observations** (Saraiya et al., 2005; J.S. Yi et al., 2008) they identify whilst using the tool. Having completed the tasks, or having felt they have thoroughly explored the community, each participant was asked to circle five characteristics of the system (drawing from the product desirability in Appendix E). Each characteristic was then used as a discussion point in a post evaluation interview, which was coupled with a set of open-ended questions. Each session lasted between one and one and a half hour. We analysed the video of each session and used content analysis, in a similar to the previous chapter, to first code and then evaluate the dept and level insights drawn from their use of the tool. We used inductive content analysis to draft an initial code book, which consisted of eleven categories. Next, using two coders, we categorised the transcribed video. This process involved two individual sessions to reach a satisfactory kappa coefficient of 0.81. We do not claim to follow North's insight based methodology to completion, although this is an approach that is gaining in popularity (Boyandin, Bertini, & Lalanne, 2012), as a larger sample is required and a longer evaluation strategy to obtain statistically significant results.

5.4.2 Results

In this section, we present the results of the video analysis and discuss the post-evaluation interviews. We were interested in understanding how participants found Petri, what their opinion was as regards specific design elements and how they considered visualisation could be assimilated into their management process. Finally, we identified and categorised any usability issues that participants encountered during their session.

5.4.2.1 Video Analysis

Each user set about the tasks differently. One participant, for example, started the first task but then abruptly abandoned it and just explored the visualisation in an attempt to find users that they considered interesting. Having participated in earlier design workshops, some participants had one goal in mind, to identify potential gurus. In two cases, the entire evaluation was driven by that goal. The other two participants followed the prescribed tasks more carefully. However, they were much less familiar with the community.

Usefulness

To assess usefulness, we wished to answer questions such as: Did users learn something about the community when using Petri? What cycles of analysis did Petri facilitate? What design element contributed to their understanding? Can Petri help a community manager make decisions about their community? What actions would a user suggest having used Petri? In accordance with pursuing a grounded evaluation, we triangulated answers from the video analysis and exit interview.

The large majority of comments that participants made during the evaluation were observational (Figure 50), these are self-evident, are of interest yet do not signify a deep level of insight. Some of the most interesting uses of the tool came when a participant sought out a user of interest. For instance, one participant, who manages the community, searched for a user because they “wanted to see how he was doing”. She had contact with him before and thought he might be “Guru material”. They found the user and then analysed his path. They said “but he seems to be getting a good bit of kudos. Maybe it is not extensive but...” Later on she chose a second user that is also of interest to her. She examined both their paths simultaneously. They commented, “I was wondering is he guru material, but now I can see he is not very active. I remember seeing him around a lot but obviously not much anymore.” In this instance, Petri helps to validate their decision as regards the potential of this user. They then commented on the second user’s path: “This girl here is quite active. I don’t know is she guru material but she seems quite motivated so maybe in the future.” Clicking on accepted solutions for these two users, she commented “Ok she (the second user) has gave [sic] a lot of replies but has no accepted solutions otherwise it is quite regular.” The ability to quickly asses how a user contributes, and how their contribution is received, enables the community manager to validate their decisions on promotion in the community. The participant indicated that this is quite important because “it always happens that you make wrong decisions (about the promotion of users) and you cannot revert back because you do not want them on your back but that is just the way it is.” Being able to reflect on the selection process and validate it qualitatively, through interaction with users, and quantitatively, through analysis with Petri, helps to act against these sorts of problems. However, it also calls into question the integration of Petri with the community’s content. In each instance, the participant would have to revert back to the community’s discussion forum and search out the individual posts to qualitatively validate their decisions as opposed to qualitatively assessing the user contribution within the context of a single interface (considered an embedded visualisation in **DC23**).

Similarly, another participant filtered for users who just joined the community this year. Finding a user that is particularly active, they said “whether she would recommend him as a guru. He looks very active across a couple of threads. Only one solution. It is doubtful that he is a guru. It depends what he is doing? If those five threads were people continuing asking questions and him answering it or if it is about him moaning about the last release and people kudoing [sic] him for it, I don’t know”. Again, this approach to analysis reflects the experience of the previous participant. While the tool can assist in the validation, qualitative analysis remains important.

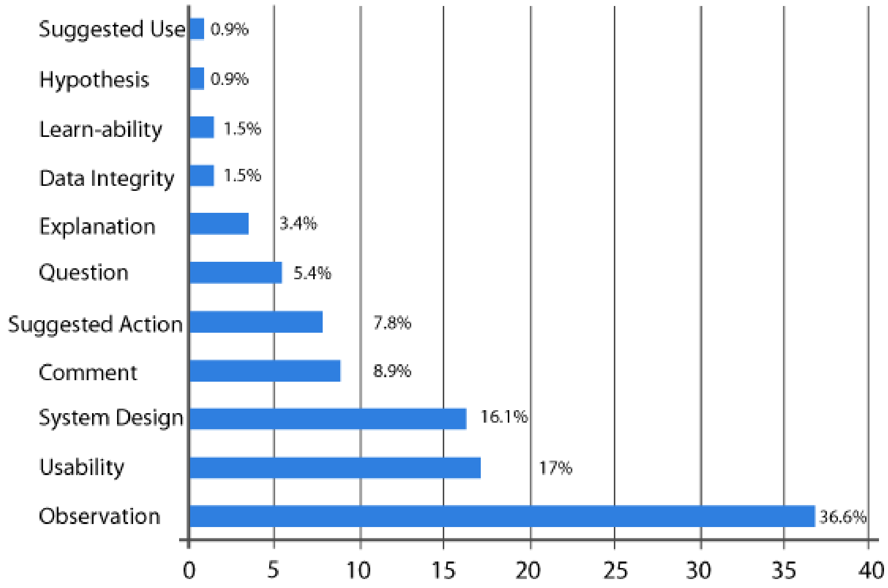


Figure 50: Content analysis of transcribed video. Categories are not mutually exclusive.

In terms of the content analysis of the video, there was very few hypothesis-driven insights (0.9%) and the majority of insights were observational (26.9%). The low number of hypotheses could be attributed to the composition of the cultures visualisation. Spatial positioning is a substantially more proficient way to support the identification of outliers, particularly pre-attentively, and outliers yield analytic insights that can result in further cycles of analysis. In addition, first degree ego centric networks are less insightful than second degree, where the communication between neighbours is observable, and altering the time-span of the visualisation could have yielded further hypotheses. Given that further iterations are required, this content analysis provides a baseline to assess the usefulness of these iterations.

Usability

Overall users found Petri usable. They found that each visualisation was legible and comprehensible. Several times, however, participants expressed some confusion with a specific aspect of a visualisation (25% of usability issues). For example, one participant had not encountered a stacked graph before and, thus, found the interactive path visualisation difficult to interpret **(DC18)**. They said that preferred the line graph was an approach they had used regularly on a personal running application. The approach to the cultures visualisation required further explanation, particularly in relation to how outliers were represented and the how each user is organise according to their distance to the medoid. The majority of usability issues (36% of usability issues) arose around the user selection process. Simple actions, such as adding or removing single users at each stage of analysis, were expected. Participants also expected a context menu when they right clicked a user's glyph. Similarly, several participants tried to hold the shift or ctrl key and then click on several glyphs in a row to select multiple users. More details, better labelling, or the ability to interrogate the actual numbers, was also identified as an issue by some participants (11%). There were some issues as regards terminology (17%) as participants found certain words and descriptions confusing (especially the word "analyse" used to select a user for analysis). Finally, some participants felt that the navigation did not accurately reflect the analytic methodology (6%) and in two cases filtering by date caused confusion (5%).

5.4.2.2 Interviews

Although the analysis helped us evaluate the usefulness of Petri, we also wanted to better understand the context in which the tool could be applied. We conducted a short exit interview with each participant, in which we drew upon the themes of usefulness, design, improvement and the process of community management. We consider these themes as core components of reflective community design. Here we briefly discuss the result of the interviews.

Usefulness

All users indicated that they found Petri to be useful. Several focused on specific design elements. One of the participants said that even in a short time that the evaluation took place, she can "say immediately now, how many gurus there are, how sparse they are. Proportionally, how much they contribute. The other thing you can tell right away is that they are not alike. Gurus are their own people; they are not cookie-cutters. The tool could be very insightful for encouraging users, and provide key indicators." Others commented on how quickly they came to terms with the interface, as it was "easy to find", "inviting" and "straightforward", suggesting that the tool was easy to learn and had a low barrier to use.

Interestingly, users had differing opinions on the approach to categorising the cultures. While one participant queried the "use-case" behind having a pragmatist culture, a second suggested that the pragmatist culture may be of most interest, as this is the body of users that they wish to engage in peer-based activity. One of the participants, who is not engaged with any one community but oversees strategy, said that Petri would give them much more "visibility" into the community and would provide grounds for validating the decision-making of her community mangers.

Design and Integration

Not many participants used the network visualisation. One participant said that they found the ego-centric network too restrictive, commenting that "it might be too focused on the individual, it does not tell me anything about whole network look like", and that they would prefer the ability to explore the whole of the network, even if that was a very large and dense graph. Another suggested that the network should be composed of the users that are selected for further analysis.

The category of lurker was discussed also. Two participants identified that it was important to have an understanding of the proportion of lurkers in the community even if that category is overly represented in the culture visualisation. Having the ability to show/hide lurkers was suggested.

Improving Petri

Some participants discussed the methodology around the application of Petri. One participant suggested that there is a clear methodology associated with the use of the tool, which involves observing shifts in cultures and then drilling into user's paths and networks. This, however, is not clear from the design and could benefit from a better description. They suggested that, at present, it looks as if these three visualisations are in "parity". Some participants also expressed dissatisfaction with the network functionality. As mentioned previously, they found it less useful than the other two visualisations and indicated that they would have preferred access to the entire network as opposed to the ego-centric visualisations provided in Petri.

Other aspects of the tool were also discussed. Embedding the visualisation **(DC23)** was proposed by several participants, whereby the user could easily traverse between Petri and the forum content whilst maintaining the context of analysis. This design consideration is discussed in Sect. 3.6.3 and enables the community manger to assess the "quality" of a user's contribution, coupling qualitative with quantitative analysis. Interaction was also raised in relation to the path visualisation. Two participants specifically asked for better "drill down" and required the ability to reduce and increase the span of analysis. Flagging employees, who are active across the entire forum, was also suggested[75].

Finally, the scrutability of the visualisation was also highlighted. While in general the participants responded positively to the interface, two participants asked to see the "actual numbers", to support their understanding of what is presented in the visualisation and to assist in their analysis.

Reflective community design at Symantec

We were interested to see whether participants thought Petri could help facilitate current community management practices at Symantec. We asked about adoption, and specifically sought to establish how Petri could fit into their current workflow. While participants noted the value of the tool, they were a little unsure of how it could be adopted. Reflecting on this, one participant commented:

> Possibly, it is very cool, forward looking and it has potential. The community manager on her day to day job, when she has ten things on her plate, typically, she will look for something to improve productivity. This is a "more big" [sic] picture tool. Would it be used on a day to day basis? I think there would be a novelty to it first but then I think it has to offer something that drives productivity or would require a very strong individual use case.

This raises the question of productivity and the use of explorative tools to support existing work practices. Although reflective community design can be undertaken at varying stage of a community's development and be supported by a range of different visualisations, for visual analysis to successfully integrate into the existing work practice there needs to be more of a focus placed on how tools can increase productivity.

Another participant suggested that community managers require a more reflective approach to the process of community management more generally, and thought Petri could add value in this context. They described it in this way:

> Community managers have very particular things to think about, i.e. their everyday list of issues to be addressed. But they have to have a more abstract view. It's like, say you were raising a child, and you have to look after it day-to-day, and you fed it and you clothed it and all those things. But, if you have a more abstract view of child development, then you have other things to think about that aren't apparent on a day to day meter. Have reflexives set in, has cognitive learning set in, has speech happened? You know, it is a different process and I don't think we have that learning for community managers to step back and say, well really, given the development curve of a community we should be here and we are not.

This same participant suggested that while the analysis of one community is interesting, comparing communities (as considered in (**DC24**) may provide much more fruitful results.

Decisions about guru selection were considered by other participants. Some participants suggested that Petri could reduce the time required to assess new candidates because you would not be "required to read all the posts and assess the user's contribution" and that the tasks could be completed in "no-time at all". One participant went on to describe how they would use the tool in practice:

[75] It is stipulated that Symantec employees use their Symantec email address.

> If I wanted to determine a new guru, and I didn't really know who, if there were three people who were potential gurus, I could use this to see how active are they and then go back to the forums to see how much quality they are providing.

They would not use the tool daily but probably "weekly" or "bi-weekly" and "it could provide you with the metrics if you want to make a case for a specific user". Another participant queried the adoption of such tools given, generally speaking, the technical knowledge of online community managers. They suggested that adoption of such advanced visual analytic tools could meet with some resistance from those less technical.

5.5 Limitations

In pursuing a human-centred approach to visualisation, we sought to include users into the design process as much as we possibly could. We drew on Munzner's nested model for design to identify points at which feedback, and thus iteration, could be realised, namely, analytic abstraction (or algorithm design), visual mapping, and interaction. While our aim was to pursue this approach in as methodical a way as possible, collaborating with a busy team, that travel widely and attend to a company's entire community strategy, presented some practical difficulties. We would have liked, for instance, to carry out additional workshops to further refine initial sketches (visual mappings) and to identify alternative, and possibly more adept, interactive strategies (addressing in particular **DC18** and **DC19**). Following the domain analysis, mapping the user requirements to different visual presentations, as discussed in (Card & Mackinlay, 1997), or using a taxonomy, as discussed in (Tory & Moller, 2004), may have help formalise this design process further. It would have been interesting, for example, to include the design considerations from Table 8 in the participative design workshops as a way to possibly communicate or indeed codify the requirements of the community management team. For the reader, a taxonomic approach could help clarify certain design decisions; while for the reviewer of a publication, such an approach could provide the grounds for better critical analysis. We consider this further in the Sect. 6.5 of the conclusion chapter. Our approach to analytic abstraction was initially presented in the third workshop, as discussed, and then later refined through discussions (both face-to-face and over email) with members of the community management team. We do not claim that each iteration, as described in this chapter, validated the approach we took, but that each iteration provided further justification for our design approach. The evaluation conducted at the end of the process helped to assess the degree of usefulness of individual desing elements as well as the overall system.

There are also some limitations as regards to the generalisability of the existing Petri implementation. While the community in question is small, given the scope of Symantec's online communities more generally, participants had asked in early workshops could this tool be easily ported to other communities? Certainly, the approach to clustering **(DC3)** is generalisable and can be reapplied to other datasets, once, that is, a degree of noise is accounted for and thus filtered out. In the case of Petri, we only included users who had posted over three times to the community. Of course, there are other issues when visualising datasets of scale. From a technical standpoint, handling a huge dataset can be problematic, particularly if the visualisation requires interaction such as pan, zoom and filter and thus cannot be pre-rendered. Approaches such as "multiscale data aggregation", as applied in Google maps and used by Elmqvist et al. for their work on Zame (Elmqvist, Do, Goodell, Henry, & Fekete, 2008), can allow the user zoom into various sections of a large dataset. But how this tool would be used in practice is an open question, particularly considering the busy schedule of the community manager. Could a community manager use this tool in regular intervals to assess how a given policy is proceeding, especially given the size of some popular online communities? Better interactive strategies would be required, and ways to quickly and intelligently reduce the dimensionality of the dataset needed **(DC19)**. We developed Petri to dovetail into the community management practice, so that the community manager could regularly asses the results of their interaction with the community. Such an approach may be much less possible with large online communities, suggesting a design that enables at-a-glance monitoring with embedded visual widgets **(DC23)**.

5.6 Discussion of Design Considerations

In general Petri was well received by the community management team. Similar to Explore.SU, as discussed in the previous chapter, as would be expected, the number of observations outweighs the number of hypotheses generated during the study (Figure 50). However, in this instance, less than 1% of the comments were hypothesis-driven, which are considered the genesis for further cycles of analysis. In this section, we consider the different design elements and suggest amendments where appropriate. As with the discussion in Sect. 4.6, we do not claim that our findings are generalisable beyond the context of the current study but that the study allows us to refine elements of the framework in Table 8. Also, any analysis should take into account the implementation of the selected design considerations.

5.6.1 Integration and Configuration

Although during the participative design workshops, the community management team expressed concern at the implementation of a coordinated visualisation, the use of a linked visualisation reduced the efficacy of the participant's visual system by requiring the user to traverse between visualisations yet retain a composite mental map. Alternatively, integrating each visualisation (potentially with a different cultures and network visualisations) into a single coordinated visualisation or an interactive dashboard can potentially capitalise on their perceptual capabilities to identify correlations between different visual elements. In considering this proposition, we suggest the following amendment to **DC22:**

Table 14: Suggested amendment to Linked visualisation

DC22 Linked Visualisation	Consider productivity and potential usage scenarios with the application of linked visualisation.

5.6.2 Visualisation 1: Cultures

The approach to the cultures visualisation included several design considerations from Table 8, most notably clustering based on user contribution **(DC3)** coupled with interactive exploration to enable analysis and support filtering **(DC19)**. Participants during the study understood the organisation and representation of the community; however they showed less interest in analysing different cultures and instead focused on identifying potential gurus in the pragmatist and apprentice culture. Two participants in the study considered the interface as more of "big picture" sort of tool as opposed to a device that could easily integrate into their existing work practice and potentially increase productivity. This is not to say the approach is unsuccessful, given the numerous ways in which clustering on contribution can be implemented and the results rendered in a visual representation, but that the participants in the study showed a preference for a more refined implementation. This could be based on the size of the community given that the community has a relatively small population and the community managers already have a good understanding to the extent of contribution. Alternatively, the need to zoom before filtering may have reduced the potential of identifying outliers, or users who have unusual attributes in relation to their position in the community **(R5),** in other cultures. However, if each user was coloured in accordance to their culture and then spatially organised as a scatter plot (in which the x-axis and y-axis represent different user attributes) then filtering could provide more insight. This is an interesting design problem to consider, given that the dataset is heavily skewed.

Nevertheless, reducing the scale of the visual search space to assist in more rapid analysis suggests an amendment to **DC19**:

Table 15: Suggested amendment to Interactive Exploration

DC19 Interactive Exploration	Consider reducing the size of the visual search space before interactive exploration.

5.6.3 Visualisation 2: Paths

The aim of the paths visualisation was to provide a compact way to analyse user contribution. Several design elements were drawn from Table 8, most notably the use of interactive animation to maintain orientation when transitioning from a line graph to a stacked graph **(DC18)**. One participant expressed some confusion with this approach, not so much by the actual animated transition but in the use of a stacked graph to represent contrasting user actions. Other participants found little difficulty with this approach but made no reference to the use of interactive animation. A problem with this design element however, is that several different visual representations, each having particular information value, are conflated into one visual widget and then interaction is used to provide access to this information. While this is a compact way to present the data, and interactive animation can potentially assist in mental map perseveration, there is also the potential for the approach to reduce the perceptual capabilities of the user. Alternatively, drawing on the small multiples design consideration **(DC25)**, each graph (reflecting user attributes over time) could be presented horizontally and each user could then be presented vertically. This configuration would enable the community manager to quickly compare and contrast activity based on perception as opposed to relying on interaction to access each representation individually. Moreover, the community manager must transition back and forth between representations to compare and contrast user contribution. This analysis suggests the following amendment to **DC20**:

Table 16: Suggested amendment to Animated Transitions

DC20 Animated Transitions	Consider the trade-off between information value and compact visual widgets.

5.6.4 Visualisation 3: Ego-centric Networks

Out of the three visual representations, the network was the least popular. Only two participants considered analysing user's network in any depth. As mentioned in the interviews (Sect. 5.4.2.2), one of the participants found the approach too restrictive in the sense that she was unable to access the entire network or, alternatively, access a section of the network using a search and discover interface **(DC10)**. The implementation could have made this approach more useful by, for instance, enabling the user to increase the size of the network and using icons to label each node **(DC11)** (as opposed to using labels on the left hand side of the interface to reference the current node). Moreover, while we sought to maintain a legible interface, and not provide a cluttered and un-readable display, we could have implemented a more compelling visual representation as discussed in Sect. 3.4.1. Readdressing **DC13** suggests the following amendment:

Table 17: Suggested amendment to Animated Transitions

DC13 Reveal Socio-temporal Dynamics	Consider representing ego-networks at a distance of two or greater.

Finally, one design consideration that was not implemented but was requested was the ability to embed the visualisation within the context of the community existing content (**DC23**). While implemented in different applications, chapter 3, this design consideration is considered of particular importance give the need for both qualitative and quantitative analysis.

5.7 Summary

This chapter investigated how to design a visual analytics system that can enable or enhance the reflective practice of online community managers at Symantec. In consultation with Symantec's online community management team, we conducted a comprehensive domain analysis and drafted a set of user requirements for a system that could support existing management practices. Based on these user requirements, we selected a set of the design considerations from Table 8 and implemented these considerations in a visualisation system called Petri. We evaluated the usefulness of these design considerations in several exploratory user studies with members from the Symantec community management team and triangulated our results with post-evaluation interviews. Participants indicated that the tool could be used periodically to assist in selecting community members for promotion or to assess the development of the community over-time but also made alternative suggestions, indicating a preference for a more confirmative tool that focused on a subset of users and was associated with the underlying community content. The need to productivity was emphasised and in light of this we considered alternative implementations and suggested amendments to several of the design considerations in Table 8. We also suggested that Table 8 could potentially assist in support a more human-centred approach to visualisation by providing an initial set of design considerations for discussion within the group. In the next chapter we present the conclusions of this thesis, review the contribution and consider directions for future work.

6

Conclusions

"Reflection is the practice of periodically stepping back to ponder the meaning to self and others in ones immediate environment about what has recently transpired."

Joseph Raelin, Public reflection as the basis for learning, 2001

6.1 Introduction

The research presented in this thesis investigated how to design visual representations that support reflective community design. Drawing on the information visualisation and visual analytics literature, we developed a set of design considerations and then, through two case-based implementations, evaluated their applicability to two community contexts. This chapter reflects on what we have accomplished and examines what potential avenues there are for further work. Sect. 6.2 outlines the objectives used to guide this research. Sect. 6.3 describes the accomplishments of this research while Sect 6.4 considers the value of these achievements in terms of our contribution. Sect 6.5 reflects on the work presented in this thesis while Sect 6.6 examines some potential avenues for future work. Sect 6.7 concludes the thesis with some closing remarks

6.2 Research Questions and Objectives

In this thesis, we addressed the following research question:

RQ: How can we design visual representations that support reflective community design?

In considering this research question, we outlined the following objectives:

RO1: Establish a basis for reflective community design.

RO2: Analyse how online communities are currently represented using visualisation technology and select a set of design considerations based on this analysis.

RO3: Evaluate the applicability of these design considerations through two independent case-based implementations with different users, different communities and different visualisation systems.

6.3 Achievements

To establish a basis for reflective community design, and thus address the first research objective (RO1), we discussed online communities in terms of socio-technical systems. It was argued that addressing online communities from this perspective introduces a range of challenges that can inhibit the design process (Sect. 2.3). Given that socio-technical systems are inherently complex and subject to continuous change, there is a need for a continuous design cycle, one that is not focused on a particular stage of community development but distributed throughout a community's lifetime. In addition, there is a need for community designers to be community members, familiar with the conventions of the community and positioned best to address a continuous design cycle. Based on this analysis, we reviewed a range of existing methodologies, such as traditional information systems development, of which online communities are a specialised sub-class, participative design and a range of approaches under the heading community design (Sect. 2.4). However, given how online communities are complex and evolving, many existing methodologies are not sufficiently capable of addressing a dynamic design space. There is either too great an emphasis on technical over social development or a reliance on individuals external to the community process. As a result, and based on the writing of Donald Schön, in Sect. 2.5 we suggest reflection, when coupled with action, provides a conceptual framework with which to address a range of complex and dynamic design situations. Research has shown that cycles of reflection and action enable the implementation and subsequent evaluation of design decisions at varying degrees of resolution. This is evident in the traditional design disciplines where designers "cast their ideas into a material form" and then use these representations to reflect on the suitability of their design decisions, to think critically about their design approach and to share their thinking with other designers. At the same time, this can also be found in organisational development where representations such as the balanced scorecard enable managers to reflect on the implementation of different organisational strategies. In each instance, it is by reflecting on action that individuals and groups refine their thinking and learn from experience. This argumentation established a basis for reflective community design, however, the need for valid and useful representational techniques remained pertinent.

To address the second objective (RO2), we analysed how online communities have been visually represented using techniques drawn from the information visualisation and visual analytics literature and selected a set of design considerations based on this analysis (chapter 3). This review looked at the literature from a range of different perspectives yet maintained a focus on novice or casual users. This involved an inquiry into how to create an engaging experience using animation (Sect. 3.5.1) and how to support discussion and analysis using a range of different collaboration mechanisms (Sect. 3.7). Although collaboration is not an essential part of visual representation, conversation is considered an important part of online community design, particularly given the emphasis on co-creation (Sect. 2.3.2). Having established a set of design considerations, we assessed how each design consideration has been evaluated to date and tabulated the results in Table 8. It was found that, of the approaches reviewed, only once was collaboration considered and the resulting implementation was never evaluated. Further, only of the four approaches reviewed include users during the design process and, in many cases, there was a preference for demonstration over evaluation. This is a particular critique of the social visualisation literature and of several approaches that include unconventional visual mappings. The resulting design considerations, presented in

Table 8, provide a design framework with which to build and validate specific implementations across a range of different community contexts.

To address the third research objective (RO3), we conduct two independent case studies in which we implement different visual analytics systems and evaluate the design decisions in an appropriate context. This approach enables us to critically assess their feasibility within the context of an application, consider alternative implementations and suggest amendments to the framework. The first case study (chapter 4) investigated how collaborative visual analytics can catalyse discussion and analysis in an instance of an online Q&A community called Super User. The case study involved an analysis and selection of design considerations from Table 8, an implementation of these design decisions in a system called Explore.SU and an evaluation of the different design elements in a live deployment with members of the Super User community. The system enabled participants to generate several interesting insights that were then shared within the Explore.SU system as well as within their own conversational spaces. Drawing on usage data and the results from exit surveys, the collaborative or social aspect of the design was particularly well received, and several of the design considerations (Table 8) were re-addressed in light of this evaluation. The second case study (chapter 5) took a human-cantered approach to design and included several participative design workshops with members of the community management team at Symantec. This study enabled the development of a set of requirements for a visual analytics system that could support the existing practice of community managers at Symantec. The requirements were mapped to a selection of design considerations drawn from Table 8, which were then implemented in a system called Petri. The results from this study were triangulated with insights gathered through post-evaluation interviews. Several design considerations that were implemented in Petri were re-evaluated in light of these results, particularly given the participant's focus on productivity and the need for less explorative toolsets.

6.4 Review of Thesis Contribution

This major contribution of this thesis is a set of design considerations that can guide the visual representation of online communities (chapter 3). Presently, the information visualisation and visual analytics literature lacks a generalisable design framework that addresses a wide range of common design problems that visualisation designers may need to address when visually representing online communities. Moreover, this set of design considerations focuses specifically on supporting novice - users less versed in analytic methodologies and visualisation techniques. Although we evaluate a subset of this framework in each of the two case studies, we envision designers and researchers refining and extending the framework through the implementation and evaluation of different design elements in different community contexts as well as extending and adapting the framework to address additional design problems.

The minor contribution of this thesis is a demonstration of the utility of this framework by analysing two community instances, implementing a set of design considerations based on this analysis and then evaluating the suitability of these design considerations in an appropriate context, i.e. with the correct domain users using a grounded evaluation strategy (chapters 4 and 5). Each case study enables us to consider our design decisions in context and to suggest ways in which to improve their implementation as well as considering how to extend and refine the overall framework.

6.5 Discussion

Although an underlying aim of this thesis is to forward our thinking on how reflection can support the continued advancement of online communities - which we consider an increasingly important social and organisational component in contemporary society - our specific focus has been on the design of visual representations for online community users. Within this context, we sought to better understand how to create visual representations that are both intuitive and engaging. To address the research question in chapter 1, we devised a set of design considerations; we outlined how a subset of this framework could be used to address two specific community instances with different design goals and different users; we developed and deployed two systems and we evaluated our design elements within context and, where appropriate, suggested amendments to Table 8. Although the design considerations provide the designer an initial set of building blocks with which to consider the visualisation design space, the mapping of the individual design elements to the community's social configuration suggests a design process that, if advanced further, could help to increase our design knowledge and reduce the potential for ineffective representation. Previous work has sought to, in varying degrees, formalise the visualisation design process. Mackinlay's early work, for example, involves automating the graphical presentation of relational data (Mackinlay, 1986), Chi's data state reference model identifies the major steps in the visual representation of data (Chi, 2000), while Munzner's nested framework supports the selection and validation of design elements independent of context or application (Meyer, Sedlmair, & Munzner, 2012). In contrast, the work presented in this thesis is context dependent and specific to the visual representation of online communities for online community users. With further work, and validation through empirical research, the design framework in chapter 3 could be reduced into a number of discrete steps, referenced or implemented as a decision tree for instance, which could be used to facilitate the analysis and selection process as undertaken in chapters 5 and 6. This approach may not only reduce the potential for ineffective design to emerge from the design process but also, potentially, encourage community users themselves to act as visualisation designers.

6.6 Future Work

This thesis has demonstrated the suitability of a framework for the visual representation of online communities. However, the work has also presented several avenues for further inquiry. Although we have initiated additional case studies to help answer fundamental questions the applicability of our design considerations and the as regards the practicality of reflective community design, further research could focus on the following challenges:

6.6.1 Rapid Iterative Development and Validation

Although in this thesis we conduct two independent case studies, the applied procedure in the second case study (chapter 5) could have benefitted from a more rapid and iterative development cycle. However, similar to co-design and drawing on element of agile software development, this requires the visualisation designer to work closely with community designers and develop "quick and easy" visual representations in response to rapid cycles of feedback. Drawing on the design framework presented in Table 8, the visualisation designer can use workbench technologies such as R or Python's matplotlib to conduct continuous formative evaluations and ascertain the suitability of their design at different points during the development cycle. Now, the designer's role resembles that of a data scientist and only after several formative evaluations, during which the utility of the different design considerations have been established and the potential for adoption enhanced, does the designer implement the system (using a D3 for instance). The idea of this approach is to continually refocus the development on the needs of the community designers and to enable the visualisation designer appreciate how the system will address the practicalities of reflective community design.

6.6.2 Towards a Pattern Language

The design considerations proposed in the thesis provide the foundation for a set of best practices that can support the development of visual representations for support reflective community design. There is scope to advance the design considerations from an initial set of propositions towards a pattern language through a socially-orientated process that would seek to develop consensus as regards their application in specific contexts. This would include a methodology for validating individual propositions, a systematic process for refining the framework and a means to reach consensus as regards amendment. Through this process, the initial set of propositions presented in Table 8, could be removed, refined, advanced and re-organised but also a methodology for their implementation could be defined.

6.6.3 Wizard or Domain Specific Language

Although this thesis proposed a design framework for the visual representation of online communities, there remains a need as illustrated in both case studies (chapters 4 and 5) for the development of actual visualisation systems. While visualisation toolsets have reduced the challenge of implementation, and the increasing prevalence of visualisation on the web has the potential to raise levels of visual literacy, the development process remains an obstacle to enabling reflective community design. One approach, which warrants further investigation, is the creation of a Wizard or Domain Specific Language (DSL) that could enable community designers rapidly generate visualisation widgets relevant to their specific needs. Using the design framework presented in Table 8 as a foundation for this approach, a simple wizard interface or DSL could lower the cost of integration, in terms of development and learnability, which in turn increases the potential for adoption. Some work has already been proposed in this area. For example, Kriplean et al.'s approach to social translucency in Wikipedia proposes a template-based DSL to assist in the generation of embedded visual widgets (Kriplean, Zachry, Borning, & McDonald, 2010) while McKeon takes a similar approach to embedding visual dashboards in Wikipedia's interface (McKeon, 2009), however there is considerable scope to broaden this application to other communities beyond Wikipedia.

6.6.4 Evaluating Reflective Community Design

Although the work suggested in Sect. 6.6.1 and 6.6.3 seeks to further validate, refine and build upon the design considerations presented in Table 8, there remains a need to evaluate reflective community design as a methodology for online community development. The case studies presented in this thesis focus on validating the design of visual analytic systems within a specific context as opposed to evaluating how reflection supported by visual representation can facilitate the development of more productive online communities. It has been argued in the online community literature that the complexity of online community design is best understood using longitudinal research methodologies (de Moor, 2007b; Schoberth, Preece, & Heinzl, 2003). Similarly, visualisation researchers have argued that longitudinal research methodologies are required to establish the utility of visual analytics systems when applied within a particular context (Shneiderman & Plaisant, 2006). Thus, the system-in-use approach, whereby the application of technology is studied in the context of use over a sustained period of time, is most suited to evaluating reflective community design. We suggest qualitative research methodologies could be of particular benefit. Virtual ethnography, as proposed by Hine (Hine, 1998) for example, could help to establish how reflective community design can practically benefit community building and community maintenance in communities with different organisational configurations.

6.7 Closing Remarks

Whether connecting across disparate geographic regions or contributing to an Open Source project, the ability for individuals to coalesce into online communities has made a significant impact on modern society. Nevertheless, online communities can be fragile and fleeting, and successful community design requires considered and continuous attention from the community members most invested in their success. This thesis proposes a set of design considerations that can assist in the creation of visualisation systems. These systems can help community designers to deepen their understanding of community dynamics but also enable them to reflect on, and indeed learn from, how different design decisions come to impact these dynamics over time. Much work is still required in refining the design framework and in conducting further case studies; however, we hope that this work contributes towards our understanding of how to develop increasingly robust online community environments.

7. Bibliography

Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Hum.-Comput. Interact.*, *15*(2), 179–203.

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, Construction (Center for Environmental Structure Series)* (p. 1171). Oxford University Press (1977).

Anderson, C. (2006). *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion.

Andrews, K. (2006). Evaluating information visualisations. In *BELIV '06 Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (pp. 1 – 5).

Archambault, D., Purchase, H., & Pinaud, B. (2010). Animation, Small Multiples, and the Effect of Mental Map Preservation in Dynamic Graphs. *IEEE transactions on visualization and computer graphics*, *17*(4), 539–552.

Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative science quarterly*, *21*(3), 363–375.

Aslak Burkhard, R., Andrienko, G., Andrienko, N., Dykes, J., Koutamanis, A., Kienreich, W., … Brodbeck, D. (2007). Visualization Summit 2007: ten research goals for 2010. *Information Visualization*, *6*(3), 169–188.

Bacon, J. (2009). *The Art of Community: Building the New Age of Participation (Theory in Practice)* (1 edition., p. 341). O'Reilly Media.

Barton, J., Emery, M., Flood, R. L., Selsky, J. W., & Wolstenholme, E. (2004). A Maturing of Systems Thinking? Evidence from Three Perspectives. *Systemic Practice and Action Research*, *17*(1), 3–36.

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., Mcdine, D., & Brooks, C. (2010). Useful Junk ? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2573–2582).

Bederson, B. (1999). Does Animation Help Users Build Mental Maps of Spatial Information ? In *INFOVIS '99 Proceedings of the 1999 IEEE Symposium on Information Visualization* (Vol. 11, p. 28).

Benedek, J., & Miner, T. (2002). Measuring Desirability: New methods for evaluating desirability in a usability lab setting. In *Proceedings of Usability Professionals Association 2003 (2002): 8-12.*

Benkler, Y. (2002). Coase's Penguin, or, Linux and The Nature of the Firm. *The Yale Law journal*, *112*(369).

Benkler, Y. (2006). *The Wealth of Networks* (p. 528). Yale University Press.

Bernstein, M. S., Ackerman, M. S., Chi, E. H., & Miller, R. C. (2011). The trouble with social computing systems research. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11* (pp. 389–398). New York, New York, USA: ACM Press.

Bertin, J. (1983a). Colour Variation. In *Semiology of Graphics: Diagrams, Networks, Maps* (p. 91). Wisconsin: The University of Wisconsin Press.

Bertin, J. (1983b). Diagrams involving two components. In *Semiology of Graphics: Diagrams, Networks, Maps* (p. 195). Wisconsin: The University of Wisconsin Press.

Bertin, J. (1983c). The Retinal Variables. In *Semiology of Graphics: Diagrams, Networks, Maps* (p. 65). Wisconsin: The University of Wisconsin Press.

Bødker, S., Grønbæk, K., & Kyng, M. (1995). Cooperative design: techniques and experiences from the Scandinavian scene. In *Human-computer interaction* (pp. 1–14).

Bonsignore, E. M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D. L., & Shneiderman, B. (2009). First Steps to Netviz Nirvana: Evaluating Social Network Analysis with NodeXL. In *2009 International Conference on Computational Science and Engineering* (pp. 332–339). IEEE.

Boyandin, I., Bertini, E., & Lalanne, D. (2012). A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small Multiples. *Computer Graphics Forum*, *31*(3pt2), 1005–1014.

Boyd, D., Ramage, D., & Donath, J. (2002). Developing legible visualizations for online social spaces. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (pp. 1060–1069). IEEE Comput. Soc.

Brooke, J. (1996). SUS - A quick and dirty usability scale Industrial usability evaluation. In *Usability Evaluation In Industry*. CRC Press.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organizational Learning*, *2*(1), 40–57.

Butler, B., Joyce, E., & Pike, J. (2008). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1101–1110). Florence, Italy: ACM.

Butler, B., Sproull, L., Kiesler, S., & Kraut, R. (2007). Community effort in online groups: Who does the work and why. In Suzanne P. Weisband (Ed.), *Leadership at a Distance* (pp. 171 – 195). Psychology Press.

Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal Of Social Psychology*, *11*(1), 13–41.

Card, S. K., & Mackinlay, J. (1997). The structure of the information visualization design space. In *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium* (pp. 92–99,). IEEE Comput. Soc.

Card, S. K., Mackinlay, J., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. (S. K. Card, J. Mackinlay, & B. Shneiderman, Eds.) (p. 712). London: Academic Press.

Carpendale, S. (2008). Evaluating information visualizations. *Information Visualization*, 19–45.

Cawthon, N., & Vande Moere, A. (2007). The Effect of Aesthetic on the Usability of Data Visualization. *2007 11th International Conference Information Visualization (IV '07)*, 637–648.

Chan, J., & Hayes, C. (2010). Decomposing Discussion Forums using User Roles. In *Proceedings of the Second Web Science Conference*. Raleigh, North Carolina.

Chen, A. (2011, April). Wikipedia Is Slowly Dying. *Gawker*. Retrieved from http://gawker.com/5827835/wikipedia-is-slowly-dying

Chen, C. (2005). Top 10 unsolved information visualization problems. *IEEE computer graphics and applications*, *25*(4), 12–6.

Chernoff, H. (1973). The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of the American Statistical Association*, *68*(342).

Chi, E. H. (2000). A taxonomy of visualization techniques using the data state reference model. *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, *94301*(Table 2), 69–75.

Chi, E. H. (2009). A position paper on ' Living Laboratories ': Rethinking Ecological Designs and Experimentation in Human Computer Interaction Outdated Evaluative Assumptions. In *Human-Computer Interaction. New Trends.* (Vol. 597–605, pp. 597–605.).

Danis, C. M., Viégas, F. B., Wattenberg, M., & Kriss, J. (2008). Your place or mine?: visualization as a community component. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 275–284). Florence, Italy: ACM.

De Moor, A. (1999). *Empowering communities: A method for the legitimate user-driven specification of network information systems*.

De Moor, A. (2007a). Using System Dynamics to Construct Design Theory on Community Information Systems. *Journal of Community Informatics*, *3*(1), 1–21.

De Moor, A. (2007b). Using System Dynamics to Construct Design Theory on Community Information Systems. *Journal of Community Informatics*, *3*(1), 1–21.

Delugach, H. S. (2007). An evaluation of the pragmatics of web-based bug tracking tools. In *Proceedings of the 2nd international conference on Pragmatic web - ICPW '07* (pp. 49–55). New York, New York, USA: ACM Press.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI EA '11 CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428).

Dibbell, J. (1993). A Rape in Cyberspace (or How an Evil Clown, a Haitian Trickster Spirit, Two Wizards, and a Cast of Dozens Turned a Database Into a Society). *The Village Voice*, 36–42.

Donath, J. (1995). Visual who: Animating the affinities and activities of an electronic community. In *ACM Multimedia 95 - Electronic Proceedings*. San Francisco, California.

Donath, J. (2002). A semantic approach to visualizing online conversations. *Commun. ACM*, *45*(4), 45–49.

Donath, J. (2007). Signals in Social Supernets. *Jorunal of Computer Mediated Communication*, *13*(1).

Donath, J., & Dragulescu, A. (2009). *Data portraits: aesthetics and algorithms*. *Media*. MIT.

Donath, J., Dragulescu, A., Zinman, A., Viégas, F., & Xiong, R. (2010). Data Portraits. *Leonardo*, *43*(4), 375–383.

Donath, J., Karahalios, K. G., & Viégas, F. B. (1999). Visualizing conversation. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences* (Vol. 32, p. 9). Hawaii: IEEE Comput. Soc.

Duncan, C. A., Eppstein, D., Goodrich, M. T., Kobourov, S. G., & Martin, N. (2010). Lombardi Drawings of Graphs. In *roc. 18th Int. Symp. on Graph Drawing (GD 2010).* (pp. 1–13).

Ellis, G., & Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (pp. 109 – 116). Venice, Italy: ACM Press.

Elmqvist, N., Do, T.-N., Goodell, H., Henry, N., & Fekete, J.-D. (2008). ZAME: Interactive Large-Scale Graph Visualization. In *Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific* (pp. 215–222). Kyoto, Japan: IEEE.

Erickson, T. (2009). Socio-Technical Design. In Brian Whitworth & A. De Moor (Eds.), *Handbook of Research on Socio-Technical Design and Social Networking Systems* (pp. 334 – 336).

Erickson, T., & Kellogg, W. A. (2000). Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, *7*(1), 59–83.

Fischer, G., Giaccardi, E., Ye, Y., Sutcliffe, A. G., & Mehandjiev, N. (2004). Meta-design: a manifesto for end-user development. *Communications of the ACM*, *47*(9).

Fischer, Gerhard, & Herrmann, T. (2011). Socio-technical systems: a meta-design perspective. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, *3*(1), 1–34.

Fisher, D. (2005). Using egocentric networks to understand communication. *Internet Computing, IEEE*, *9*(5), 20–28.

Fisher, D., & Dourish, P. (2004). Social and temporal structures in everyday collaboration. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, 551–558.

Fisher, D., Smith, M., & Welser, H. T. (2006). You Are Who You Talk To: Detecting Roles in Usenet Newsgroups. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2006. HICSS '06.* (p. 59).

Flynn, P. J. (2000). Data Clustering : A Review. *ACM Computing Surveys*, *31*(3), 264–323.

Flyvbjerg, B. (2011). Case study. In *The Sage Handbook of Qualitative Research, 4th Edition* (pp. 301–316). Thousand Oaks, CA: Sage.

Forte, A., & Bruckman, A. (2008). Scaling Consensus : Increasing Decentralization in Wikipedia Governance Andrea Forte and Amy Bruckman. In *Proceedings of the 41st Hawaii International Conference on System Sciences (HICSS)* (pp. 15–17). IEEE Computer Society.

Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure*, *1*. doi:10.1016/j.dss.2010.02.001

Furtado, A., & Andrade, N. (2013). Contributor profiles, their dynamics, and their importance in five q&a sites. In *Proceedings of the 2013 …* (pp. 1237–1252).

Gazan, R. (2009). When online communities become self-aware. In *System Sciences, 2009. HICSS'09. 42nd Hawaii …* (pp. 1–10).

Ghoniem, M., Fekete, J.-D., & Castagliola, P. (2004). A Comparison of the Readability of Graphs Using Node-Link and Matrix-Based Representations. In *IEEE Symposium on Information Visualization* (pp. 17–24). IEEE.

Gilbert, E., & Karahalios, K. (2007). CodeSaw: A social visualization of distributed software development. In *INTERACT'07 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction - Volume Part II* (pp. 303–316).

Gilbert, E., & Karahalios, K. (2009). Using social visualization to motivate social production. *IEEE Transactions on Multimedia - Special section on communities and media*, *11*(3), 413–421.

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). A Conceptual and Operational Definition of 'Social Roles in Online Community. In *Hawaii International Conference on System Sciences* (Vol. vol, pp. 1–11). Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. *Internet Research*, *5*, 19–22.

Goldspink, C., & Kay, R. (2007). Social emergence: Distinguishing reflexive and non-reflexive modes. In *… Symposium: Emergent Agents and Socialities: Social*

Gruzd, A., & Haythornthwaite, C. (2008). The Analysis of Online Communities using Interactive Content-based Social Networks. In *In Proceedings of the American Society for Information Science and Technology (ASIS&T) Conference (October 2008),* (pp. 523–527).

Gurzick, D., & Lutters, W. (2009). Towards a design theory for online communities. In *The 4th International Conference on Design*. Philadelphia.

Hagel, J., & Armstrong, A. G. (1997). *Net Gain: Expanding Markets Through Virtual Communities* (p. 235). Harvard Business Press.

Haythornthwaite, C. (2008). Crowds and Communities: Light and Heavyweight Models of Peer Production. In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS)* (pp. 1 – 10). Hawaii: IEEE.

Haythornthwaite, C., & Gruzd, A. (2007). A Noun Phrase Analysis Tool for Mining Online Community Conversations. In *Proceedings of the Third Communities and Technology* (pp. 67 – 86). Michigan State University, East Lansing, Michigan, USA.

Heer, Jeffery, & Boyd, D. (2005). Vizster: visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (pp. 32–39). Minneapolis, MN, USA.: IEEE.

Heer, Jeffery, Viégas, F. B., & Wattenberg, M. (2009). Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM - Rural engineering development*, *52*(1), 87–97.

Heer, Jeffrey. (2008). *Supporting Asynchronous Collaboration for Interactive Visualization*. University of California, Berkeley.

Heer, Jeffrey, & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization*, *7*(1), 49–62. 7

Heer, Jeffrey, Bostock, M., & Ogievetsky, V. (2010). Visualisation: A Tour through the Visualization Zoo A survey of powerful visualization techniques , from the obvious to the obscure. *ACM Queue*, *8*(5), 20.

Heer, Jeffrey, Ham, F. Van, Carpendale, S., Weaver, C., & P. (2008). Creation and collaboration: Engaging new audiences for information visualization. *Visualization*, 92–133.

Heer, Jeffrey, & Hellerstein, J. (2009). Data Visualization and Social Data Analysis. In *Proceedings of the VLDB Endowment* (pp. 1656–1657).

Heer, Jeffrey, & Robertson, G. (2007). Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics*, *13*(6), 1240–7.

Henry, N., & Fekete, J.-D. (2006). MatrixExplorer: a dual-representation system to explore social networks. *IEEE transactions on visualization and computer graphics*, *12*(5), 677–84.

Henry, N., Fekete, J.-D., & McGuffin, M. J. (2007). NodeTrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics*, *13*(6), 1302–9.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, *28*(1), 75–105.

Hine, C. (1998). Virtual Ethnography. Bristol, UK.

Hornbæk, K., & Hertzum, M. (2011). The notion of overview in information visualization. *International Journal of Human-Computer Studies*, *69*(7), 509–525.

Howard, R. (1993). *The virtual community: homesteading on the electronic frontier*. *Cambridge (MA)* (p. 447). The MIT Press; revised edition edition (November 1, 2000).

Høyrup, S. (2004). Reflection as a core process in organisational learning. *Journal of Workplace Learning*, *16*(8), 442–454.

Huberman, B. a., Romero, D. M., & Wu, F. (2008). Social Networks that Matter: Twitter Under the Microscope. *First Monday*, *14*(1 - 5), 1–9.

Husson, F., Le, S., & Pages, J. (2010). *Exploratory Multivariate Analysis by Example Using R* (p. 240). CRC Press.

Iriberri, A., & Leroy, G. (2009). A life-cycle perspective on online community success. *ACM Comput. Surv.*, *41*(2), 1–29.

Isenberg, P., Zuk, T., & Collins, C. (2008). Grounded evaluation of information visualizations. In *BELIV '08 Proceedings of the 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization* (pp. 1–8). Florence, Italy.

Jahnke, I. (2009). Socio-technical communities: From informal to formal? In Brian Whitworth & A. De Moor (Eds.), *Handbook of Research on SocioTechnical Design and Social Networking* (pp. 763–777).

Karahalios, K. G., & Viégas, F. B. (2006). Social visualization: exploring text, audio, and video interaction. *CHI'06 extended abstracts on Human factors …*, 1667–1670.

Keim, D. A. (2000). Designing pixel-oriented visualization techniques: theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, *6*(1), 59–78. 1

Keim, D. A., Schneidewind, J., & Sips, M. (2006). Scalable pixel based visual data exploration. *of the 1st first visual information*, *4370*(4370), 12–24.

Keim, D., Mansmann, F., & Thomas, J. (2010). Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations …*, *11*(2), 5–8.

Kensing, F., Simonsen, J., & Bodker, K. (1996). MUST: A method for participatory design. In *Proceedings of the Fourth Biennial Conference on Participatory Design* (pp. 129–140).

Kerr, B., & Cheng, L. T. (2006). Growing bloom: design of a visualization of project evolution. In *CHI EA '06 CHI '06 Extended Abstracts on Human Factors in Computing* (pp. 93–98). ACM.

Kim, A. J. (2000). *Community Building on the Web : Secret Strategies for Successful Online Communities* (p. 380). Peachpit Press.

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: Conflict and Coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (p. 453). New York, New York, USA: ACM Press.

Kollock, P., & Smith, M. (1994). Managing the Virtual Commons: Cooperation and Conflict in Computer Communitie. *Journal of Commputer Mediated Communication (JCMC)*.

Kriplean, T., Beschastnikh, I., & McDonald, D. W. (2008). Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In *CSCW '08 Proceedings of the 2008 ACM conference on Computer supported cooperative work* (Vol. 8, pp. 47–56).

Kriplean, T., Zachry, M., Borning, A., & McDonald, D. W. (2010). Enabling Social Translucence for Wikipedia. *CSCW*.

Krishnamoorthy, S., & North, C. (2005). Learnability of Interactive Coordinated-View Visualizations. *Ninth International Conference on Information Visualisation (IV'05)*, 306–311. doi:10.1109/IV.2005.70

Larman, C., & Basili, V. R. (2003). Iterative and incremental developments. a brief history. *Computer*, *36*(6), 47–56.

Latour, B., Mauguin, P., & Teil, G. (1992). A Note on Socio-Technical Graphs. *Social Studies of Science*, *22*(1), 33–57.

Lau, A., & Vande Moere, A. (2007). Towards a Model of Information Aesthetics in Information Visualization. In *2007 11th International Conference Information Visualization (IV '07)* (pp. 87–92). Ieee.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge University Press.

Lessig, L. (2006). Creative Commons License. In *CODE 2.0* (pp. 110 – 111). Basic Books.

Liu, K., & Dix, A. (1997). norm governed agents in cscw. In *Paper presented at the 1st International Workshop on Computational Semiotics,*. PARIS - LA DEFENSE - FRANCE.

Ljungberg, J., & Holm, P. (1996). Speech acts on trial. *Scandinavin Journal of Information systems*, *8*(1), 29 – 52.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, *5*(April 1986), 110–141.

Maclean, D., & Hangal, S. (2009). You Didn ' t Tell Me That ! Visualizing the Hidden Attributes of Online Health Communities. *Cancer*, 2–5.

Marie, T., Card, S., & Mackinlay, J. (2009). Visualization Viewpoints Defining Insight for Visual Analytics. *Ieee Computer Graphics And Applications*, 0–3.

Marty, P. F. (2005). Factors Influencing the Co-Evolution of Computer-Mediated Collaborative Practices and Systems: A Museum Case Study . *Journal of Computer-Mediated Communication (JCMC)*, *10*(4).

Masli, M., Priedhorsky, R., & Terveen, L. (2011). Task specialization in social production communities: The case of geographic volunteer work. In *Proc. ICWSM* (pp. 217–224).

Mazza, Ricardo. (2009). Introduction. In *Introduction to Information Visualisation* (p. 11). Springer.

Mazza, Riccardo. (2009). A graphical tool for monitoring the usage of modules in course management systems. *Pixelization Paradigm*, 164–172.

Mazza, Riccardo, & Berré, A. (2007). Focus group methodology for evaluating information visualization techniques and tools. In *Information Visualization, 2007. IV'07. 11th International Conference* (pp. 74–80). IEEE.

McKeon, M. (2009). Harnessing the Information Ecosystem with Wiki-based Visualization Dashboards. *Visualization and Computer Graphics, IEEE Transactions on*, *15*(6), 1081–1088.

Meyer, M., Sedlmair, M., & Munzner, T. (2012). The Four-Level Nested Model Revisited: Blocks and Guidelines. In *BELIV 2012: Beyond Time and Errors - Novel Evaluation Methods for Visualizationond Time and Errors - Novel Evaluation Methods for Visualization*.

Modritscher, F., & Taferner, W. (2012). Visualization of Wiki-Based Collaboration through Two-Mode Network Patterns. In *2012 IEEE 12th International Conference on Advanced Learning Technologies* (pp. 658–659). IEEE.

Monk, A., & Howard, S. (1998). The Rich Picture: A Tool for Reasoning About Work Context. *Interactions*, *5*(2), 21–30.

Morzy, M. (2009). On Mining and Social Role Discovery in Internet Forums. In *2009 International Workshop on Social Informatics* (pp. 74–79). Warsaw, Poland: IEEE.

Mumford, E. (2006). The story of socio-technical design: reflections on its successes, failures and potential. *Information Systems Journal*, *16*(4), 317–342.

Munzner, T. (2000). *Interactive visualization of large graphs and networks*. Stanford University.

Munzner, T. (2009a). Visualization. In *Fundamentals of Computer Graphics, Third Edition* (pp. 675–707).

Munzner, T. (2009b). A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, *15*(6), 921–8.

North, C. (2006). Toward measuring visualization insight. *IEEE computer graphics and applications*, *26*(3), 6–9.

North, C., & Shneiderman, B. (2000). Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies*, *53*(5), 715–739.

O'Day, V. L., Bobrow, D. G., & Shirley, M. (1996). The social-technical design circle. In *ACM conference on Computer Supported Cooperative Work* (pp. 160–169). Boston, Massachusetts, United States: ACM.

O'Day, V. L., Bobrow, D. G., & Shirley, M. (1998). Network Community Design: A Social-Technical Design Circle. *Computer Supported Cooperative Work (CSCW)*, *7*(3 - 4), 315–337.

O'Keffe, P. (2008). *Managing Online Forums: Everything You Need to Know to Create and Run Successful Community Discussion Boards* (p. 320). AMACOM.

O'Reilly, T. (2005). What Is Web 2.0. Retrieved from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

Ostrom, E. (2000). Collective Action and the Evolution of Social Noms. *The Journal of Economic Perspectives*, *14*(3), 137–158.

Parr, C. (2013, March). Not Staying the Course. *Inside Higher Ed*.

Perer, A. (2006). Making sense of social networks. *CHI '06 extended abstracts on Human factors in computing systems - CHI '06*, 1779.

Perer, A., Guy, I., Uziel, E., Ronen, I., & Jacovi, M. (2011). Visual social network analytics for relationship discovery in the enterprise. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 71–79). Ieee.

Perer, A., Guy, I., Uziel, E., Ronen, I., & Jacovi, M. (2012). The Longitudinal Use of SaNDVis: Visual Social Network Analytics in the Enterprise. *IEEE transactions on visualization and computer graphics*, *19*(7), 1095–108.

Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, 265.

Perer, A., & Smith, M. (2006). Contrasting portraits of email practices: visual approaches to reflection and analysis. In *Proceedings of the working conference on Advanced visual interfaces - AVI '06* (pp. 389–395).

Perry, E., & Donath, J. (2004). Anthropomorphic visualization: a new approach for depicting participants in online spaces. In *CHI EA '04 CHI '04 Extended Abstracts on Human Factors in Computing* (pp. 1115–1118). Vienna, Austria: ACM.

Pirolli, P., Wollny, E., & Suh, B. (2009). So you know you're getting the best possible information: a tool that increases Wikipedia credibility. In *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1505–1508).

Pisano, G. P., & Verganti, R. (2008). Which Kind of Collaboration Is Right for You? *Harvard Business Review*, *86*(12), 78–86.

Plaisant, C., Milash, B., Rose, A., Widoff, S., & Shneiderman, B. (1996). LifeLines: visualizing personal histories. *Proceedings of the SIGCHI conference on Human factors in computing systems. ACM.*

Pousman, Z., Rouzati, H., & Stasko, J. (2008). Imprint, a community visualization of printer data: designing for open-ended engagement on sustainability. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 13–16).

Pousman, Z., Stasko, J., & Mateas, M. (2007). Casual information visualization: depictions of data in everyday life. *IEEE transactions on visualization and computer graphics*, *13*(6), 1145–52.

Preece, J. (2000a). *Online Communities: Designing Usability and Supporting Sociability* (p. 464 pages). Wiley; 1 edition..

Preece, J. (2000b). *Online Communities: Designing Usability and Supporting Sociability* (p. 464 pages). Wiley; 1 edition.

Preece, J. (2001). Online communities: Usability, sociabilty, theory and methods. In *Frontiers of Human-Centred Computing, Online Communities and Virtual Environments.* (Vol. 1, pp. 263–277).

Preece, J. (2004). Etiquette, empathy and trust in communities of practice: Stepping-stones to social capital. *Journal of Universal Computer Science*, *10*(3), 194–202.

Preece, J., Maloney-krichmar, D., & Abras, C. (2003). History and emergence of online communities. In *Encyclopedia of Community* (pp. 1–11).

Preece, J., & Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, *1*(1), 13–32.

Raymond, E. (1999). *The Cathedral and the BazaarThe Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary* (p. 241). O'Reilly media.

Reingold, T., & Fruchterman, E. (1991). Graph Drawing by Force-Directed Placement. *Software, Practice and Experience*, *21*(11), 1129–1164.

Ribarsky, W. (2009). Toward effective insight management in visual analytics systems. *2009 IEEE Pacific Visualization Symposium*, 49–56.

Riche, N., Lee, B., & Chevalier, F. (2010). iChase: Supporting exploration and awareness of editing activities on wikipedia. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 59–66).

Roberts, J. C. (2007). State of the Art: Coordinated & Multiple Views in Exploratory Visualization. *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, (Cmv), 61–71.

Robertson, G., Fernandez, R., Fisher, D., Lee, B., & Stasko, J. (2008). Effectiveness of animation in trend visualization. *IEEE transactions on visualization and computer graphics*, *14*(6), 1325–32.

Russell, D. M., Furnas, G., Stefik, M., Card, S. K., & Pirolli, P. (1993). The Cost Structure of Sensemaking. *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*,

Sack, W. (2000a). Discourse diagrams: interface design for very large-scale conversations. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* (Vol. 00, p. 10). IEEE Comput. Soc.

Sack, W. (2000b). Conversation map: a content-based Usenet newsgroup browser. In *Proceedings of the 5th international conference on Intelligent user interfaces* (Vol. 2, pp. 233–240). ACM New York, NY, USA.

Saraiya, P., North, C., & Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics*, *11*(4), 443–56.

Scacchi, W. (2004). Socio-Technical Design. In *The Encyclopedia of Human-Computer Interaction* (pp. 656–659). Berkshire Publishing Group.

Schlimmer, J. C., & Granger, R. H. (1986). Beyond incremental processing: Tracking concept drift. *Proceedings of the Fifth National Conference on Artificial Intelligence, 502--507. San Francisco*.

Schoberth, T., Preece, J., & Heinzl, A. (2003). Online communities: a longitudinal analysis of communication activities*., 2003. Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 216–225.

Schön, D. (1992). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems*, *3*(3), 131 –147.

Scott, J. (1991). *Social Network Analysis: A Handbook* (p. 207). Sage.

Scott, J. (2000). The Development of social Network Analysis. In *Social Network Analysis* (Third., pp. 11 – 38). Sage.

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (pp. 336–343). IEEE Comput. Soc. Press.

Shneiderman, B. (2008). Extreme visualization: squeezing a billion records into a million pixels. *Proceedings of the 2008 ACM SIGMOD international conference on Management of data ACM*.

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *BELIV '06 Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (pp. 1–7). The Pennsylvania State University: ACM.

Simonite, T. (2013, October). The Decline of Wikipedia. *MIT Technology Review*. Retrieved from http://www.technologyreview.com/featuredstory/520446/the-decline-of-wikipedia/

Skupin, A. (1996). Spatial metaphors for visualizing very large data archives. *GIS LIS-INTERNATIONAL*.

Smith, M., Hansen, D. L., & Gleave, E. (2009). Analyzing Enterprise Social Media Networks. *2009 International Conference on Computational Science and Engineering*, 705–710.

Smith, M., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., … Gleave, E. (2009). Analyzing (social media) networks with NodeXL. *Proceedings of the fourth international conference on Communities and technologies - C&T '09*, 255.

Smuc, M., Mayr, E., & Lammarsch, T. (2008). Visualizations at First Sight: Do Insights Require Training? In *HCI and Usability for …* (pp. 261–280). Springer.

Stein, K., Wegener, R., & Schlieder, C. (2010). Pixel-Oriented Visualization of Change in Social Networks. *Advances in Social Networks*.

Suchman, L. (1993). Do categories have politics? the language/action perspective reconsidered. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work* (pp. 1 – 14). Milan, Italy: Kluwer Academic Publishers.

Suh, B., Chi, E. H., Kittur, A., & Pendleton, B. A. (2008). Lifting the veil: improving accountability and social transparency in Wikipedia with wikidashboard. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1037–1040). Florence, Italy: ACM.

Suh, B., Chi, E. H., Pendleton, B. A., & Kittur, A. (2007). Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations. In *2007 IEEE Symposium on Visual Analytics Science and Technology* (pp. 163–170). Ieee.

Suh, B., Convertino, G., Chi, E. H., & Pirolli, P. (2009). The singularity is not near: slowing growth of Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (Vol. 1, p. 8). ACM.

Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything* (p. 368). Atlantic Books.

Thomas, J. J., & Cook, K. A. (2005). Illuminating the path: The research and development agenda for visual analytics. *IEEE Computer Society*.

Tory, M., & Moller, T. (2004). Rethinking Visualization: A High-Level Taxonomy. In *IEEE Symposium on Information Visualization* (pp. 151–158). IEEE.

Toulmin, S. (2003). *The Uses of Argument* (Updated Ed., p. 262). Cambridge University Press.

Trier, M. (2005). IT-supported visualization of knowledge community structures. *Proceedings of the System Sciences, 2005. HICSS'05. (pp.* 1–9).

Truex, D. P., Baskerville, R., & Klein, H. (1999). Growing systems in emergent organizations. *Communications of the ACM*, *42*(8), 123.

Tufte, E. R. (1987). *The Visual Display of Quantitative Information*.

Tufte, E. R. (1990a). Micro/Macro readings. In *Envisioning Information* (pp. 37 – 53).

Tufte, E. R. (1990b). Small Multiples. In *Envisioning Information* (pp. 67 – 81).

Tukey, J. (1980). We need both exploratory and confirmatory. *The American Statistician*.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing Usenet: Mapping Computer-Mediated Collective Action. *Jorunal of Computer Mediated Communication*, *10*(4).

Tversky, B., Bauer, J., Betrancourt, M., Europe, A. De, & St-martin, M. (2002). Animation : can it facilitate ? *International Journal of Human-Computer Studies*, *57*(4), 247–262.

Tzagarakis, M., Karousos, N., Karacapilidis, N., & Nousia, D. (2009). Unleashing Argumentation Support Systems on the Web: The case of CoPe_it! Athens, Greece.

Van Ham, F., Schulz, H., & Dimicco, J. M. (2009). Honeycomb : Visual Analysis of Large Scale Social Networks. In *INTERACT '09 Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (pp. 429–442). Uppsala, Sweden.

Vande Moere, A., & Hill, D. (2012). Designing for the Situated and Public Visualization of Urban Data. *Journal of Urban Technology*, *19*(2), 37–41.

Vande Moere, A., & Purchase, H. (2011). On the role of design in information visualization. *Information Visualization*, *10*(4), 356–371.

Viégas, F. B., Boyd, D., Nguyen, D. H., Potter, J., & Donath, J. (2004). Digital artifacts for remembering and storytelling: posthistory and social network fragments. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the* (Vol. 00, pp. 109–118). IEEE.

Viégas, F. B., & Donath, J. (2004). Social network visualization: Can we go beyond the graph. In *Workshop on Social Networks, CSCW* (Vol. 4, pp. 6–10). Chicago, USA.

Viégas, F. B., Golder, S., & Donath, J. (2006). Visualizing email content: portraying relationships from conversational histories. In*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 979–988.

Viégas, F. B., & Smith, M. (2004). Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspaces. In *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS)* (p. 40109.2).

Viégas, F. B., & Wattenberg, M. (2007). Artistic data visualization: Beyond visual analytics. *Online Communities and Social Computing*.

Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 575–582). Vienna, Austria: ACM.

Viégas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualization with Wordle. *IEEE transactions on visualization and computer graphics*, *15*(6), 1137–44.

Viégas, F. B., Wattenberg, M., Kriss, J., & Ham, F. Van. (2007). Talk Before You Type: Coordination in Wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* (p. 78). Hawaii.

Viégas, F. B., Wattenberg, M., van Ham, F., Kriss, J., & McKeon, M. (2007). ManyEyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics*, *13*(6), 1121–8.

Wagner, C., Rowe, M., Strohmaier, M., & Alani, H. (2012). Ignorance isn't Bliss: An Empirical Analysis of Attention Patterns in Online Communities. In *IEEE International Conference on Social Computing (2012) .* (pp. 101 – 110). Amsterdam, The Netherlands.

Wang, W., Wang, H., Dai, G., & Wang, H. (2006). Visualization of large hierarchical data by circle packing. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06* (pp. 517 – 520). New York, New York, USA: ACM Press.

Ware, C. (1999). Fondation for a science of data visualisation. In *Information Visualsiation, Percpetion for Design* (pp. 1 – 27).

Wattenberg, M. (2005). Baby names, visualization, and social data analysis. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 1–7.

Wattenberg, M., & Millen, D. (2003). Conversation thumbnails for large-scale discussions. *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*, 742.

Wattenberg, M., Viégas, F. B., & Hollenbach, K. (2007). Visualizing activity on wikipedia with chromograms. In *INTERACT'07 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction - Volume Part II* (pp. 272–287). Rio de Janeiro, Brazil: Springer.

Weber, S. (2005). The early history of Open Source. In *The Success of Open Source* (pp. 20 – 54). Harvard University Press.

Welser, H. T., Gleave, E., Fisher, D., & Smith, M. (2008). Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure*, *8*.

Wenger, E. (1999). *Communities of practice: learning, meaning, and identity* (p. 336). Cambridge University Press; New Ed edition.

Whitworth, B., & de Moor, A. (2002). Legitimate by design: towards trusted virtual community environments. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (Vol. 00, pp. 2831–2842). IEEE Comput. Soc.

Willett, W., & Heer, J. (2011). CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing System.* (pp. 3131–3140).

Winograd, T. (1987). A language/action perspective on the design of cooperative work. *Hum.-Comput. Interact.*, *3*(1), 3–30.

Xiong, R., & Donath, J. (1999). PeopleGarden: creating data portraits for users. In *Proceedings of the 12th annual ACM symposium on User interface software and technology* (pp. 37 – 44). Asheville, North Carolina, United States: ACM.

Xu, A., Chen, J., Matthews, T., Muller, M., & Badenes, H. (2013). CommunityCompare: Visually comparing communities for online community leaders in the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 523–532).

Yi, J.S., Kang, Y., Stasko, J. T., & Jacko, J. A. (2008). Understanding and characterizing insights: how do people gain insights using information visualization? In *BELIV '08 Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization* (pp. 1–6). Florence, Italy.: ACM.

Yi, Ji Soo, Kang, Y. A., Stasko, J., & Jacko, J. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, *13*(6), 1224–31.

Zhongbao, K., & Changshui, Z. (2003). Reply networks on a bulletin board system. *Physical review. E, Statistical, nonlinear, and soft matter physics*, *67*(3 Pt 2).

Zhou, M. X., & Feiner, S. K. (1998). Visual task characterization for automated visual discourse synthesis. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98*, 392–399.

Zhu, B. (2002). Visualizing a computer mediated communication (CMC) process to facilitate knowledge management. *Conference on Human Factors in Computing Systems*, 562–563.

Zhu, B., & Chen, H. (2008). Communication-Garden System: Visualizing a computer-mediated communication process. *Decision Support Systems*, *45*(4), 778–794.

Appendix A SUS Survey for Case Study 1

System Usability Scale

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Appendix B Short Exit Survey for Case Study 1

This survey is designed to gauge participant attitude to Explore.SU and the method of study. In keeping with the ethics guidelines for Trinity College, School of Statistics and Computing both surveys and all questions were optional.

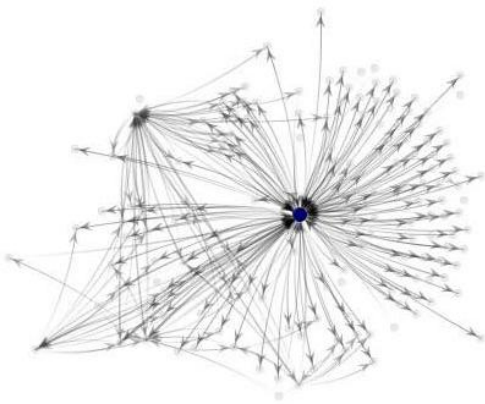
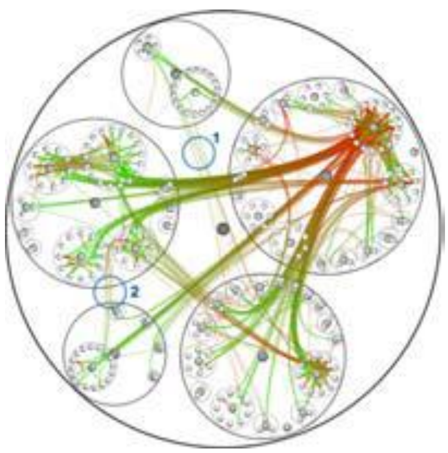
Experience of Explore.SU	
1.	Do you have any comments or suggestions in regards to the visualisation tool?
2.	Do you have any comments in regards to the process of this study?
3.	Did you enjoy using the system? (**Likert Scale 1 – 5**)
4.	Did you learn something interesting while using the system? (**Likert Scale 1 – 5**)
Community Membership Information	
5.	How long have you been a member of SuperUser.com?
6.	How often do you visit the site?
7.	Do you contribute to meta.superuser.com? (**Yes / No)**
8.	What is your SuperUser.com user id?
9.	What is your current SuperUser.com reputation?
10.	Can we contact you in the future? (**Yes / No)**
11.	If yes, please submit an email address with which we may contact you?

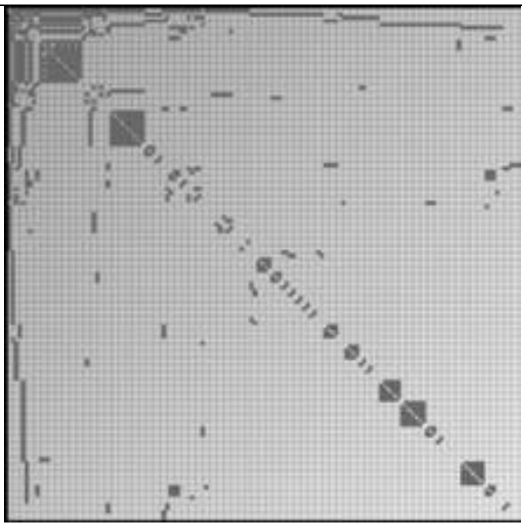
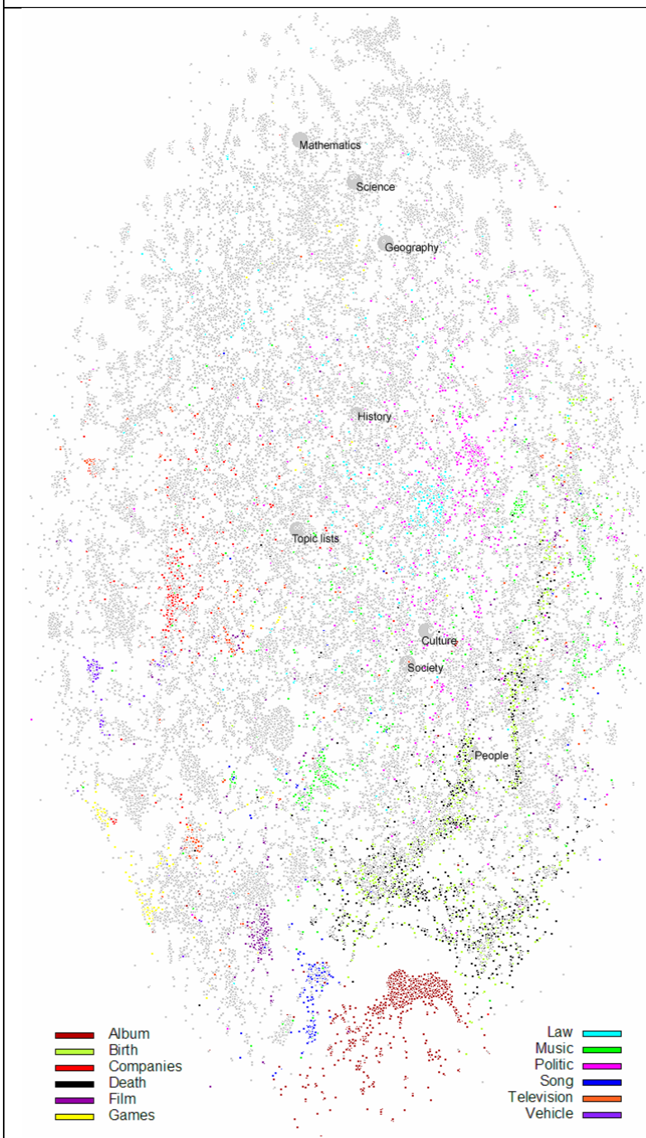
Appendix C Visualisation Information Sheet for Case Study 2

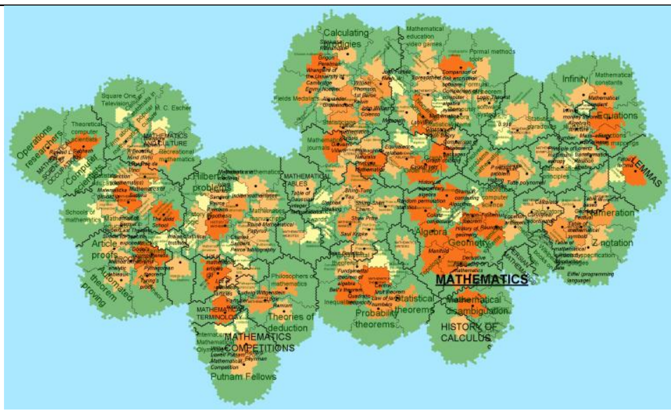
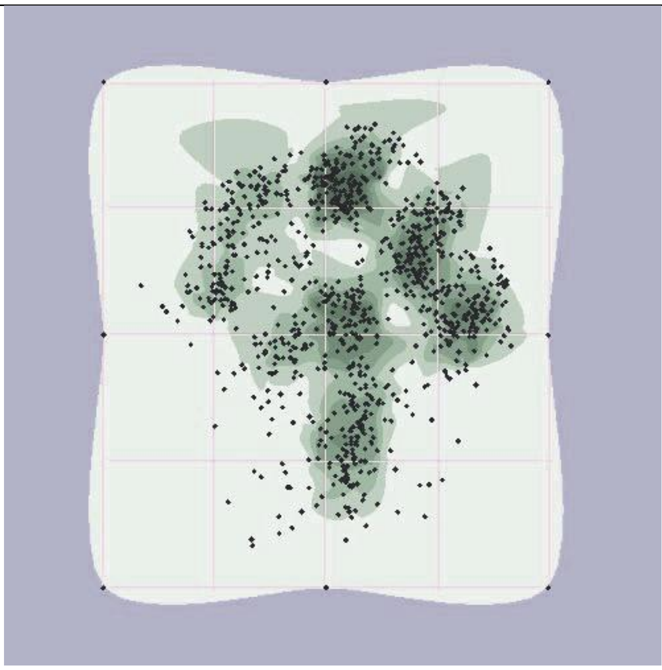
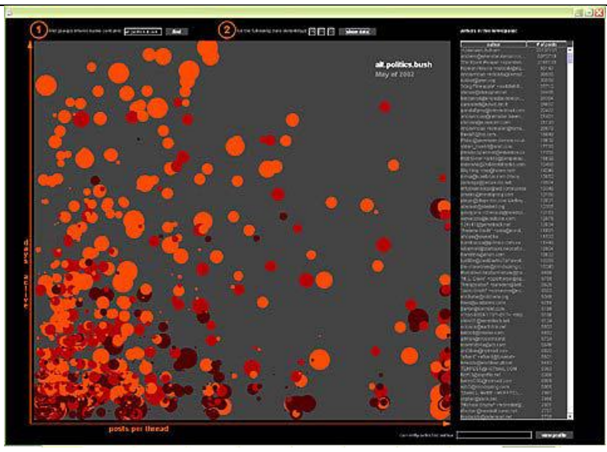
The document provides a short description of methods for visualising communities. It is divided into two parts. The first presents some examples of how to visualise an entire community. The second provides some ways to visualise individual user activity. In the second part, each visualisation was chosen as a possible way to compare a user's activity as a timeline or a path. The aim of the document is to provide some examples that can be used as discussion points in the second design workshop. Some of the examples are a little odd, or a little dated, but provided a new or interesting approach to visualisation when created.

Community Overview

The community overview attempts to present an entire community in a meaningful way.

Visualisation	Description
	This is a standard graph visualisation. The centre node is the dominant user and the links that extend from that node are connections with other users. In this case, the connections are posts sent from that node.
	In this graphic, nodes are bundled together into clusters (clustered by similarity) and the connections between the nodes are represented by the red and green lines. All the clusters are then bounded together by additional grey circles. This shows the interaction between groups (or cultures) of similar users.

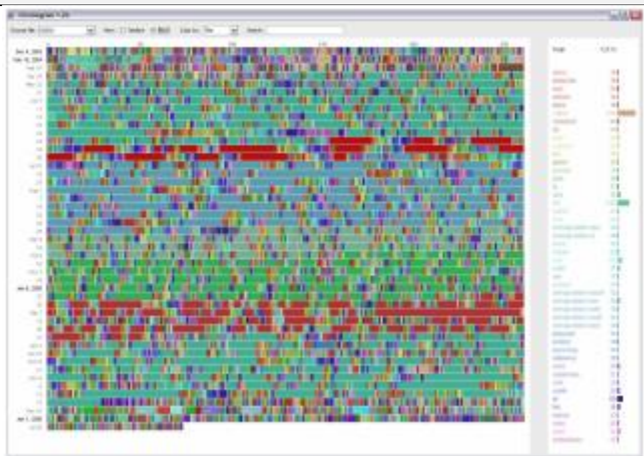
<table>
<tr>
<td></td>
<td>This is a matrix view that is good for representing large graphs that change over time. It is not as intuitive as graph-based approaches, as people tend to be less familiar with this approach. All the users are represented on the X-axis and also on the Y-axis. If a user has a connection with another user, then the intersection between those two users is coloured grey.</td>
</tr>
<tr>
<td>
</td>
<td>This visualisation is a map of categories for an entire Wiki. The basic idea is to automatically relate categories by similarity, for example football is similar to sport etc, and then map them onto a two dimensional space. The idea of a clustered map is intuitive to users and allows them navigate a wiki based on a graphical, geographic representation.</td>
</tr>
</table>

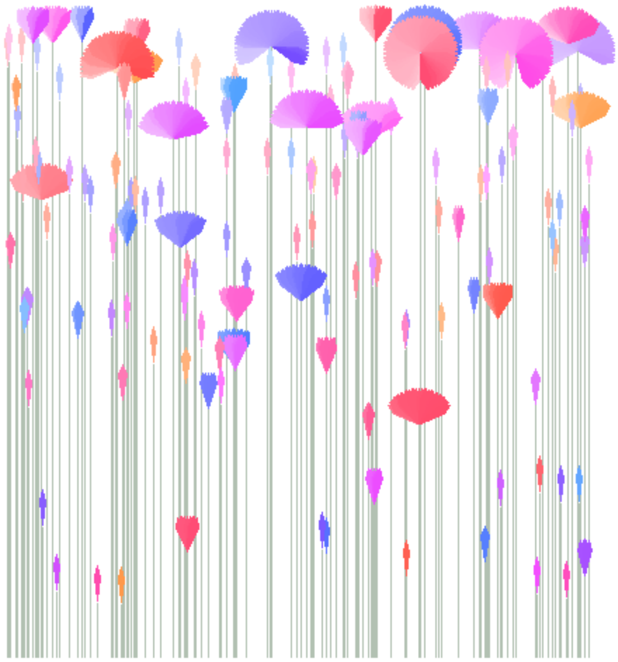
	This takes the map metaphor a little further but is based on a similar premise to the previous visualisation. Categories are related by similarity - football is liked golf etc - are clustered and then mapped.
	This is essentially a set of data that has been clustered and then mapped to a 2D space. This is standard enough approach to visualisation that takes advantage of our familiarity with maps. The clusters are highlighted by increasing the intensity of the green shading. There are a thousand points on this visualisation. I think it is very similar to what Jason proposed in the initial workshop, particularly if we were to add connections between the points (similar to the first and second visualisations). I think aesthetically, this is very pleasing also and feels more intuitive than traditional graphs.
	This visualisation is called Newsgroup Crowds. It is an advanced scatter plot in which users are characterised by their posting patterns. Using this visualisation, the creators identified a number of conversational roles that users performed in the UseNet online community. For example, debaters are highly active and respond to threads started by others that have sent a large number of messages. "Balanced" Conversationalist starts the same

	number of threads as they respond to. Pollinator responds to threads started by others etc.
	This graphic is called Bloom. It is a visualisation of user activity in an open source developer community. It presents the activity of four project roles (similar to the idea of cultures) – coder, coder-commenter, commenter-coder and commenter. This classification was predetermined by the developers as it roughly corresponds to how people behave in the project. It is basically a timeline that is wrapped in a circle. Moving along the timeline at the top moves the user around the circle – the small bubbles illustrate interactivity between users at that point in time. While this looks nice, I am unsure how effective an approach to visualisation it is.

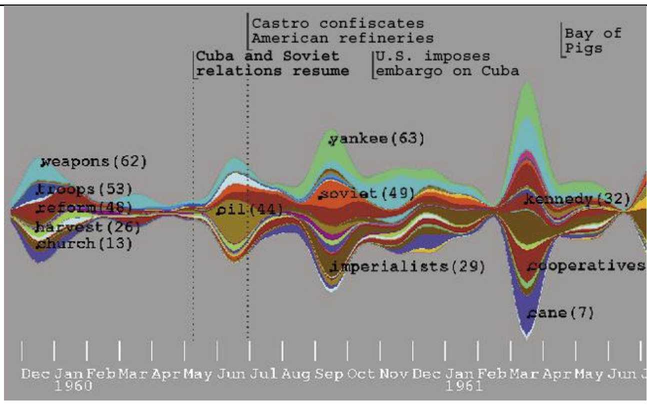
Individual Users (User path or Timeline)

The aim of this view is to enable the comparison of different user paths – their actions (plotted over time) that has enabled them to reach a certain standing in the community (for example Guru).

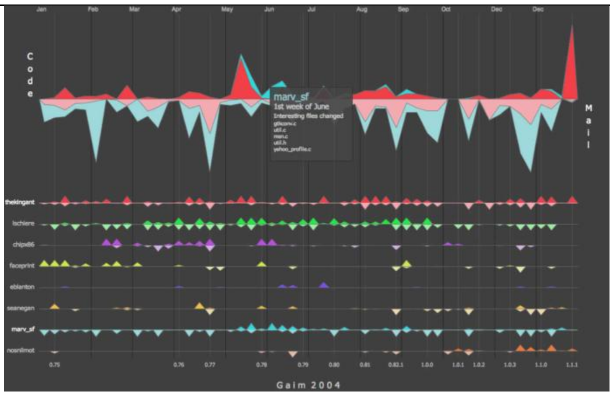
Visualisation	**Description**
	This visualisation is called a Chromogram. It represents editor activity in Wikipedia. The type of edit is represented by colour – red is a new introduction, blue is the addition of copyright or green is revert to a previous edit – and sequences of edits are represented on the X-axis. This allows the user to quickly observe and compare patterns of editor activity over time. Two patterns were observed: editors reacting to vandalism by reverting and article to a prior state and editors performing bursts of intense edit work that involved very minor edits.



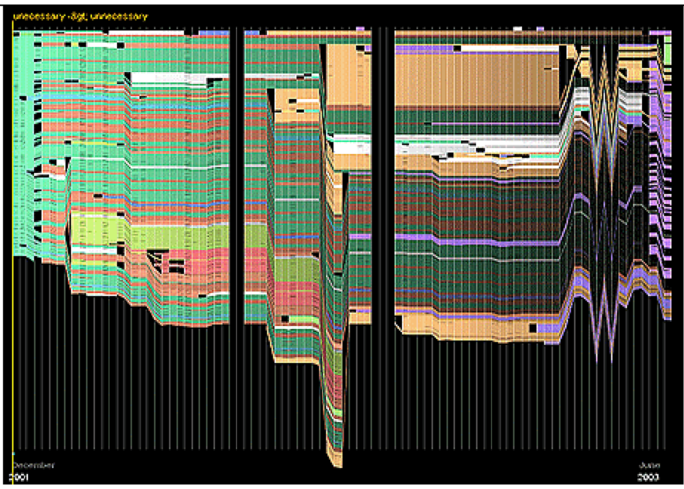
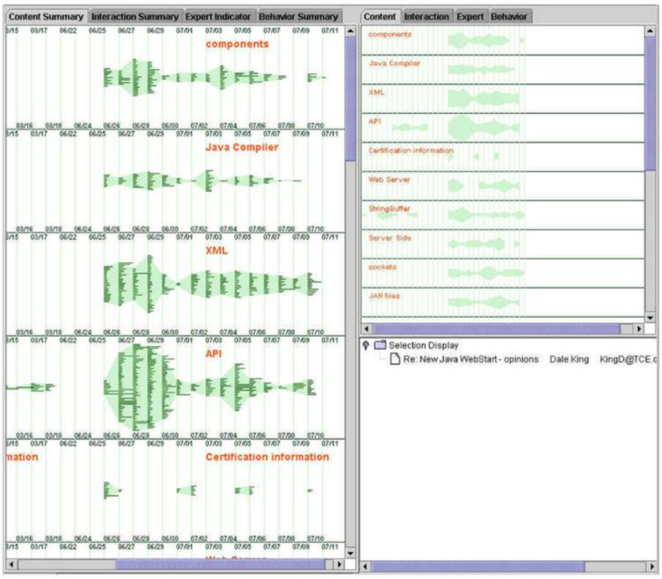
This is a more radical visualisation that uses the metaphor of a garden to visualise communities. Each user is represented by a flower. The height of the flower represents how long a user is in the community. Colour is used to denote the time of a post and the petals change over time as a user's participation increases or decreases in the community. The span of the petal is used to show the amount of replies a user receives. While this appears a little crazy the principles are interesting. Basically, visually encode some sort of user behaviour and then plot all those users on a single visualisation. Users that behave the same will appear similar in the visualisation.



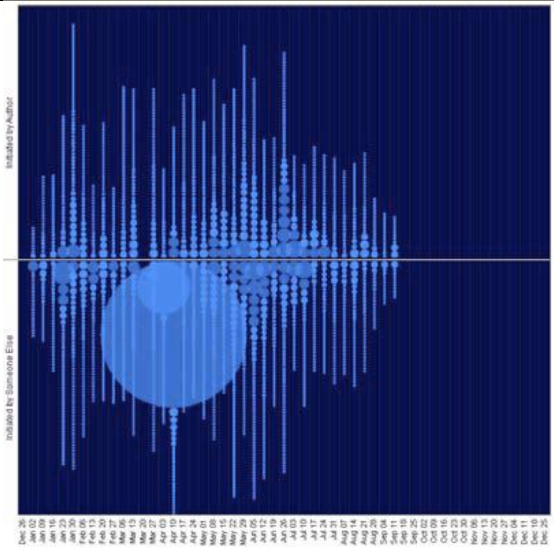
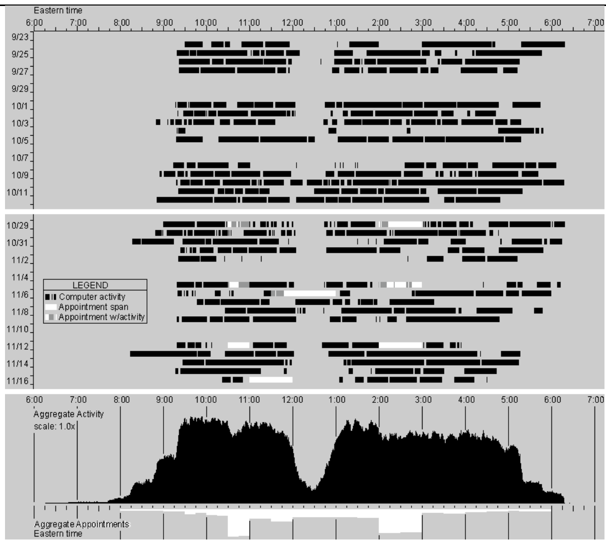
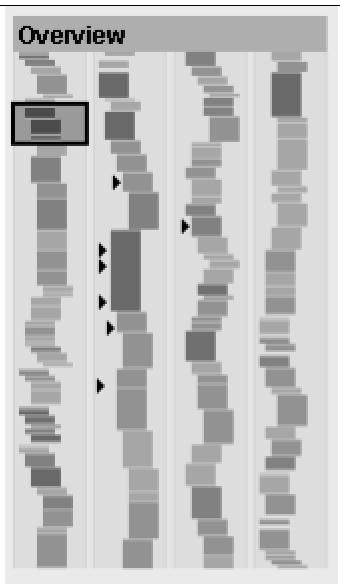
This visualisation, called theme river, is not so much about visualising communities but does illustrate changes in themes over time – this may help us to come up with a way to compare the paths of different users.

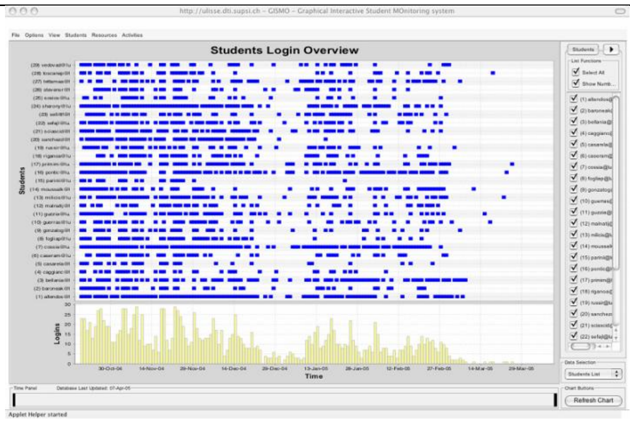
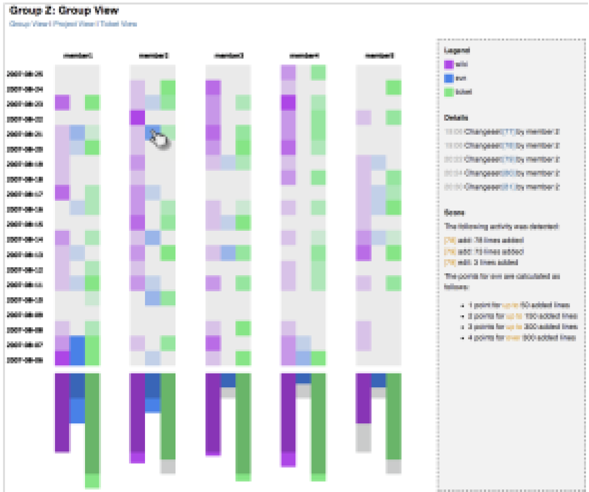
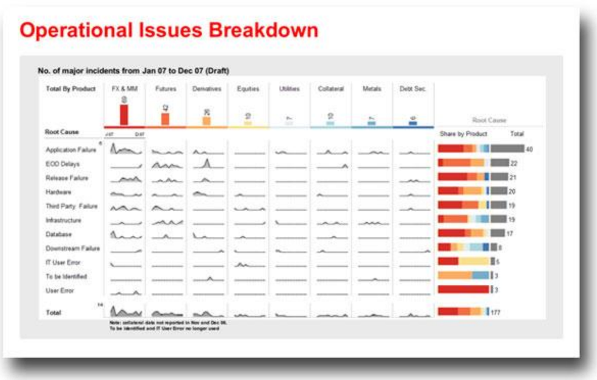


This visualisation, called CodeSaw, focuses on user timelines in an open source developer community. Each user is represented by a horizontal timeline. The peaks above the line are contributions to the codebase and the troughs below the line are emails sent by that user. This allows us to quickly assess contribution rates versus communication rates. This is similar to the visualisation Authorlines (the blue visualisation below), which presents two contrasting activities that, when juxtaposed, help to illustrate how a user contributes to a community.

<table>
<tr>
<td></td>
<td>This visualisation is called History Flow. It visualises the edit activity of a Wikipedia page. Each colour represents an edit by a particular user. Edits that occur close together (in time) are represented beside one another in the visualisation. Thus, the zig-zag patterns represent people who are editing the article in close succession (which can illustrate conflict between users). I think this approach to visualisation is similar to the Chromogram visualisation above and can show the activity of a user in relation the activity of other users.</td>
</tr>
<tr>
<td>

Description of the Display panel:
The x-axis represents time.
Categories generated by the SOM are laid out vertically.
Each green line represents one message.
The vertical thickness of each subtopic indicates its activity on a particular day.
The length in the x-dimension of each subtopic = time duration of that subtopic.</td>
<td>This visualisation, called Communication Garden, shows the activity of messages on several message boards over time. Time is represented on the X-axis and each small dark green, horizontal dash represents a message for that day. The general shape of all the dashes together are highlighted by light green. This provides the messages on each board with a shape - as can be seen in the top right panel of the graphic on the right. This approach could be used to represent and then compare and contrast the path of each user. So, each user's path will have a general shape (the Guru shape for example) and we can compare the paths of each user by their shape.</td>
</tr>
</table>

	This visualisation is called Authorlines. The idea is that threads that are initiated by a user in a community are represented above the line and threads replied to by a user are represented below the line. The number of communications is represented by size. So, the large blue bubble is a thread that a user is contributing a lot to but has not initiated.
	This graphic is called an Actogram. It visualises a person's work activity over time. Dates are on the X-axis and times are on the Y-axis. The graphic is broken into three horizontal panels. The top panel visualises the user's computer activity. The middle horizontal panel overlays their computer activity with their appointments and the last panel is an aggregate view of their activity. I included this example as it provides another way to compare time-orientated data.
	This is a thumbnail (or small) view of a conversation in a discussion board. It illustrates how discussion boards (given reply threads etc) typically develop – in this stratified way. The overview and zoom allows the user to focus on a particular part of the conversation. In this graphic, each thread could be considered as a user's path and the presentation allows us to compare and contrast how each user's path changes over time. So, in this example the four threads could be viewed as four users and the colours and indents could be either actions or attributes of those users – posts, replies, amount of time in the community

	etc.
	This is a very simple visualisation that places users on the horizontal (X) axis and the date on the vertical (Y) axis. Each blue dot indicates when a user logged into the system. So, quite simply, a manager can see who and when a user logs into the system. At the bottom, a histogram illustrates overall activity. This is a simple example but does illustrate a way to compare user's activity with a single screen.
	This visualisation compares the activity of five users of a small community. The date is listed on the left and their activity on either a wiki (purple), a forum (blue) or a messaging client (green) is highlighted for that day. The intensity of the colour indicates the level of activity. The bottom is an aggregation of all the users' activity. This visualisation enables the comparison of user activity, however, it is very simple and not very nuanced.
	Another option is based on the idea of small multiples – several small visualisations presented in close proximity, each representing a different characteristic of a certain entity. So, for example, in the image on the left, each vertical column (from red to blue) could be a user and the horizontal rows could be a certain characteristic of that user - number of questions, number of answers, peer-support etc - all represented over time. This approach may help to compare and contrast different users.

Appendix D Open-ended tasks for Case study 2

The following simple, open-ended tasks were proposed to help the community management team familiarise themselves with the Petri system. E

1. Locate the users who contribute the most to the community.
   a. Analyse each user.
   b. Compare and contrast their contribution over time.

2. Locate a user from each of the four cultures (Pragmatist, Peer, Apprentice and Guru):
   a. Analyse each user.
   b. Compare and contrast their contribution over time.
   c. Compare their kudos and reply networks. Comment on what you observe.

Question: Are there any actions you would take having completed this task?

3. Identify users of interest using the follow approach:
   a. Consider a culture of interest.
   b. Filter based on relevant attributes.
   c. Select users of interest.
   d. Compare their contribution over time.
   e. Compare their kudos and reply-to networks. Comment on what you observe.

Question: Are there any actions you would take having completed your analysis?

4. Locate users who joined the community this year.

Question: Any potential gurus in this selection?

5. Locate a Guru who has not signed in for a while.
   a. Compare and contrast their contribution.

Appendix E Product Desirability Survey for Case Study 2

Product reaction survey as outlined in (Benedek & Miner, 2002). The product reaction survey product provides a quick and easy way to gather user reactions to a product. The results are not generalisable across all potential users of the system but can be used to quickly obtain feedback to a product or desires. The tool was used during the evaluation of Petri to help initialise the interviews with participants.

Accessible	Desirable	Gets in the way	Patronizing	Stressful
Appealing	Easy to use	Hard to use	Personal	Time-consuming
Attractive	Efficient	High quality	Predictable	Time-saving
Busy	Empowering	Inconsistent	Relevant	Too technical
Collaborative	Exciting	Intimidating	Reliable	Trustworthy
Complex	Familiar	Inviting	Rigid	Uncontrollable
Comprehensive	Fast	Motivating	Simplistic	Unconventional
Confusing	Flexible	Not valuable	Slow	Unpredictable
Connected	Fresh	Organized	Sophisticated	Usable

Consistent	Frustrating	Overbearing	Stimulating	Useful
Customizable	Fun	Overwhelming	Straight Forward	Valuable