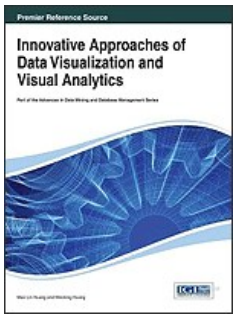

Innovative Approaches of Data Visualization and Visual Analytics

by Mao Lin Huang and Weidong Huang (eds)

Chapter 17: Community Management Matters—Advanced Visual Analytics for Online Community Managers

John McAuley

Alex O'Connor,
Centre for Next Generation Localization, Trinity College
Ireland

Dave Lewis,
Centre for Next Generation Localization, Trinity College
Ireland

ABSTRACT

Online communities provide technical support for organisations on a range of products and services. These communities are managed by dedicated online community managers who nurture and help the community grow. While visual analytics are increasingly used to support a range of data-intensive management processes, similar techniques have not been adopted into the community management field. Although relevant tools exist, the majority is developed in the lab, without conducting a domain analysis or eliciting user requirements, or is designed to support more general analytic tasks. In this chapter, the authors describe a case study in which we design, develop, and evaluate a visual analytics application with the help of Symantec's online community management team. The authors suggest that the approach and the resulting application, called Petri, is an important step to promoting online community management as a strategic and data-driven process.

1 INTRODUCTION

Over the last number of years, large organisations have recognised the benefit in hosting online communities that promote products and services and assist in the provision of technical support. These communities help to reduce the organisation's overhead, as it is the community members, and not the organisation's technical support department, that handle questions from other product and service users. The role of the online community manager has emerged as integral to this strategy. It is their job to ensure that the community develops in a productive fashion, that anti-social behaviour is kept to a minimum and that pro-active users are encouraged and in the promoted to positions of influence in the community. At the same time, the community manager must encourage "peering", whereby members of the community, and not the organisation's staff, answer questions from other community members. Active or prolific users are often then rewarded with affiliation programmes, access to new products or services, or in some cases, jobs, contracts and new positions.

The growing trend in strategic and operational management is to make use of advanced analytic technologies that provide insight into large volumes of data and thus support a more informed or data-centric approach to decision-making (Lohr, 2012). The online community manager is in an ideal position to take advantage of this trend. Community interactions are stored in online databases, which are accessible, generally, by public or private APIs. Moreover, the online community and social analytics literature is replete with new applications and technologies that could help the community manager to identify users of potential (H. Welser, Cosley, Kossinets, & Lin, 2011), spot points of conflict (Kittur & Kraut, 2010) and deepen their understanding of community dynamics (Suh, Chi, Kittur, & Pendleton, 2008). As yet, however, these tools have not been widely adopted by online community managers and either remain as part of a growing academic literature or are employed by researchers or data analysts. Furthermore, very few, if any, of these applications are designed from the perspective of the online community manager. Instead, researchers tend to develop a novel community metric and asses its utility across multiple datasets (Chan & Hayes, 2010; Wagner, Rowe, Strohmaier, & Alani, 2012) or design a new visualisation and use it to discover and thus explain certain community phenomenon (Wattenberg, Viégas, & Hollenbach, 2007).

Our position is that a human-centred design approach is required to shift visual analytics from the research community into the practice of online community management. In supporting this position, we conducted a case study with Symantec's online community management team in Dublin, Ireland. We drew on Munzner's nested model for design and validation to guide and to an extent formalise the design process (Munzner, 2009). The model consists of four nested layers, domain analysis, data/operation abstraction design, encoding/interaction technique design and finally algorithm design, and provides a rigorous way to construct and validate human-centred visualisations. While Munzner suggests validation over evaluation, we tend to adopt both methods, applying validation at certain points during the design process but then completing a final evaluation of the application with members of the community team (five in total). We tried to apply this model in as faithful a way as possible but were restricted by the amount of access we had to the team. We held five sessions in total with the community management team. The first two workshops were used to gather requirements and to sketch out some initial designs (Domain Analysis and Operation Design). The third workshop helped to validate the initial mock-ups and reject some alternative designs (Encoding/Interaction Design). In the fourth workshop, we identified the features that were used for clustering the community into meaningful cultures, which is discussed in more detail later in this chapter (Algorithm design). These cultures were then validated in discussions with the community manger that is responsible for that particular community. The fifth and final session took the form of an explorative evaluation, following the qualitative methodology proposed by Dix (Ellis & Dix, 2006). At each point, we aimed to validate the application with the intended user community, a more grounded approach that avoids drawing subjects from the lab (Isenberg, Zuk, & Collins, 2008). In total, the research was conducted over a ten month period, from January 2012 to October 2012. Next, we briefly describe the community, on which the visual analytics application was based, and then provide an outline of the chapter.

1.1 The Community

The community is a non-English speaking technical support forum that is hosted by Symantec. The forum was started in December 2009. At the time of writing, the community had approximately 7000 users. The aim of the forum is to provide technical support on a range of products and services. Thus, the majority of activity is goal driven–users seek answers to specific technical questions. The forum is hosted by an external software provider. Their service is typical of online Q&A communities. To ask a question, a user initiates a thread on a relevant board. To answer, users submit replies. If the reply answers the question to a sufficient degree, the first user, who originally asked the question, will (or at least should) accept the reply as an answer. Other users who experience similar technical problems can draw on the answer, while also having access to the chain of reasoning (replies by other users) that contributed to the answer along the way. Users can give "kudos" to other users in the community. Kudos is a way of rewarding a user for a particular action, whether this involves asking an interesting question or submitting a well crafted answer, and helps to raise the profile of effective contributors. The community is managed by a dedicated online community manger. Her role is to facilitate the community, and to answer questions that are not addressed by other members of the community.

1.2 Chapter Outline

The chapter is outlined as follows: First, we discuss background in which we focus on how online community visualisation has been addressed to date, outlining the advantages and disadvantages of each. We pay particular attention to the target audience of each approach; who, for example, was the visualisation designed to support and in what context was the visualisation to be applied. The context of the application advances the discussion from describing information with graphics to the application of the visualisation in specific organisational contexts. Next, we present a case study, which is divided into three parts: domain analysis, design and evaluation. In the domain analysis, we report on three workshops that were held with the community management team. The aim of the first workshop was to develop a better understanding of the process of community management, as a set of operational and strategic directives, and to understand the requirements that the team had in terms of a visual analytics application. The second and third workshops were used to gather feedback on initial designs. Next, we present Petri. We justify our design decisions with reference to the requirements, and conclude with an evaluation of the tool.

2 BACKGROUND

While there are a substantial number of applications and research prototypes available that can be used to visualise social media and online community data, none of these are developed to specifically address the process of online community management. Existing approaches fall generally within one of two categories: Activity-based visualisation, which is based on attribute or statistical data, and network-based visualisation, which is based on communication or connection data. Under network-based visualisation, there are further sub-categories, including tools that have Graphical User Interfaces (GUI) or tools that are based on programming environments or Application Programming Interfaces (API)(Henry & Fekete, 2006). The second sub-category is highly configurable yet complex and requires a substantial level of in-depth expert knowledge or expertise. In this section, we briefly analyse each category. We consider the end-user and how appropriately the application supports end-user tasks.

2.1 Activity-Based Visualisation

Activity-based visualisation presents the community as either definable groups or repeated patterns of user activity. The advantage of this approach is that it draws upon the user's perceptual capabilities to isolate and associate significant patterns and related features. This is in contrast to network-based visualisations, which generally rely on mathematical algorithms such as spring or force-directed layouts, to position the nodes and arcs in 2D space and thus improve graph readability.

Newsgroup crowds and Authorlines are two early examples of activity-based visualisation that enabled the identification of social roles in the Usenet online community (such as "Answer person," "Pollinator," or "Debater") (Viégas & Smith, 2004). While this work made a significant contribution to online community research, there are some limitations to approach to visualisation. Newsgroup crowds is a multivariate scatter plot that, while effective, is restricted to conveying four variables at a time, size, colour and coordinates. A fifth variable, shape, could be added but can serve to clutter the resulting visualisation. While this may be sufficient, there is much more information collected about a user that could improve the granularity of the resulting representation. Also, despite the use of transparency, occlusion remains an issue, as users with a similar profile can overlap and make identification and selection difficult. Finally, from the perspective of this work, the visualisations were developed to support exploration and discovery by casual users and were not designed from the perspective of the online community manager.

Other work by Viégas et al. found that visualising personal email archives prompted reflection and storytelling in casual users (Viégas, Boyd, Nguyen, Potter, & Donath, 2004). While the domain is different, salient aspects of the approach remain relevant for online community datasets. Users can, for instance, explore their email exchange or social interactions from a temporal perspective. However, the visualisations were not developed to support external analysis but instead were designed to promote personal reflection. As a result, the visual composition is more expressive and has less fidelity than expected from an analytic application. For instance, occlusion remains problematic and the use of animation is not based on increasing the level of insight but instead aims to emphasise the ebb and flow of personal communication.

HistoryFlow (Viégas, Wattenberg, & Dave, 2004; Viégas, Wattenberg, Kriss, & Ham, 2007) and Chromograms (Wattenberg et al., 2007) are two novel visualisations that were designed to explore Wikipedia. While the first exposes the edit activity of users on Wikipedia pages, the second exposes the edit activity of Wikipedia editors (described as Wikipedians by [Panciera & Halfaker, 2009]). These visualisations were used by the authors to support inductive analysis (Kimmerle, Moskaliuk, Harrer, & Cress, 2010), so that the authors could explore and then explain, posit hypotheses about or draw conclusions from Wikipedia as a large scale socio-technical phenomenon. Each visualisation design is specific to this purpose and is, thus, removed from the needs of a community manger or other community stakeholders. Having said this, their contribution is significant and the Chromogram, as a novel visual construct, has been adopted by other community researchers (Masli, Priedhorsky, & Terveen, 2011).

In contrast, Comtella (Vassileva & Sun; 2007) and Codesaw (Gilbert & Karahalios; 2007, 2009) are two social visualisations that were

designed to support specific user groups. The aim of Comtella is to motivate users and increase participation rates in a peer-based sharing community, while Codesaw was designed to improve team coordination in Open Source Software (OSS) teams. While the work on Comtella has provided evidence to suggest that participation rates can be increased through the introduction of social visualisations, from an analytic perspective, the approach is quite limited. The visualisation, which draws heavily on Erickson's design principles (Erickson, 2003), is based on an evocative metaphor, described by the author "as a summer night's sky", and only one or possibly two visual variables are used when communicating activity to the user (size and colour). Also, as with any visualisation that lies beyond the experience of the user, the authors found that the interface needed some explanation. In this respect, a scatter plot, as proposed by Viégas previously, could be more useful, particularly when approached from an analytic perspective. The design of Codesaw, on the other hand, is of more interest, as the authors decided upon a user-centred design approach and consulted the intended user group from the outset. The resulting artefact is quite effective at illustrating the contrasting work practice of an OSS development team – the communications patterns versus the code-based contributions. However, Codesaw is designed for a small community (8 in total) and there is no way to select and, thus, analyse user behaviour from a larger online community. Naturally, this presents a problem when wishing to address communities of scale.

Finally, some researchers have created "legible" and "intuitive" visualisations or "data portraits" that convey, in a qualitative sense, the rich social features of users and communities. This work includes People Garden (Xiong & Donath, 1999), Loom (Donath, Karahalios, & Viégas, 1999), Loom2 (Boyd, Ramage, & Donath, 2002) and anthropomorphic visualisations (Perry & Donath, 2004). While certainly interesting, the integrity of the visualisation is always questionable given that the designer attributes additional meaning with often evocative metaphors. The result is more impressionistic than realistic, and may, as a consequence, not serve the purposes of the community manager (which is to comprehend in as faithful a way as possible the behaviour of their community). Also, each approach was designed to support exploration and discovery by casual users and there was little consultation with user groups or communities during the design process.

2.2 Network-Based Visualisation

There are a number of work-bench-type applications, such as R, NetworkX with Gephi and GraphViz, currently available that can be used to analyse and visualise network data. Even the most user-friendly of these applications, however, are methodologically-driven (requiring the use of "recipes" to get started) and are designed to support the explorative analysis of network data, often from the perspective of a data scientist or professional analyst. In addition, users require some programming knowledge to parse and assemble the graphs, which can lie beyond the skill-set of the regular online community manager. While applications, such as Orion (Heer & Perer, 2011) and NodeXL (Bonsignore et al., 2009; Smith et al., 2009), help to mitigate these concerns, there still remains the need for the user to understand how to apply the appropriate analytic methodology. Certainly, NodeXL, based on Microsoft's Excel software, is a step in the right direction, in terms of broadening a user community; however it remains a comprehensive analytic tool. While programming is not required, an understanding of social network analysis is; as is an understanding of how to methodically approach the analytic process. For a novice, getting to grips with the application involves a period of trial and error. And while it has been shown that students can quickly understand and thus apply NodeXL to analyse social media data, the demonstration was undertaken within a tutorial setting with the aid of an experienced tutor. How an online community manager would adjust to the introduction of NodeXL, thus, remains an open question.

Less complex network applications, such as Vizster (Heer & Boyd, 2005) and ICTA (Gruzd & Haythornthwaite, 2008), provide users with access to online community data as visualised social networks. While certainly more intuitive, Vizster was designed to support "playful discovery" of community networks, from the perspective of the casual end-user, and was not designed from the perspective of the online community manager. Having said this, Vizster's approach to the interactive exploration of community networks remains quite innovative. ICTA (Internet Community Text Analyzer), on the other hand, is a content-based approach to the analysis and visualisation of online community data. Unlike other network visualisations, ICTA extracts usernames from community posts and then creates a social network based on those names and the connections between the posts. The authors argue that the use of a name carries increased significance and thus provides another way to evaluate the community's network. Both Vizster and ICTA only support exploration, however, and while explorative analysis is important for discovery, from our analysis we learned that community managers require the ability to shift from explorative to confirmative analysis, when validating the choice of super-user for example. Shifting modes of analysis requires precise information delivered to the user in different formats at different times.

Researchers have also investigated the use of matrix-based representations to visualise social connections in groups and communities. While less popular than node-link diagrams, matrices have some advantages, namely that they not only show the connections between users but also show incidents where there are no connections between users, thus, the user is privy to how groups are communicating and also not communicating. Matrices can also be used to illustrate temporal shifts more effectively than node-link diagrams due to the nodes having a fixed position in space. Van Ham et al. Developed Honeycomb, a scalable matrix-based visualisation, to address the communications of large organisations (such as IBM). They found meaningful communications across the entire organisation, using varying degrees of resolution, such as country and business division. One advantage of this approach is that it provides the user with a single reference model, unlike a composite of metrics and visualisation interfaces like NodeXL. However, the authors focused on emphasising the benefits of matrices over node-link diagrams, as opposed to addressing how matrices could be adopted by non-expert users, such as online community managers, in other analytic scenarios. In contrast, Henry's work with social science researchers sought to establish how matrices could be used to support their work practice (Henry & Fekete, 2006). Taking a user-centred design approach, Henry characterised the domain of the social science researchers, drafted up a set of requirements and then designed and developed a matrix-based visualisation environment to address their needs. While not specific to the domain of online community management, the approach to design, coupled with the goal of applying visualisation to support an existing user-community, illustrates how to create meaningful visualisation environments for non-expert users (i.e. non-professional analysts).

Finally, the use of network or matrix-based visualisation places an overriding emphasis on the connection, a view that was first established by the sociologist Georg Simmel (Marin & Wellman, 2010), over the other characteristics of the social group. This view has gained considerable currency in online community research, which is, to some degree, driven by the ease of access to electronic records and early publications, such as (Danyel Fisher, Smith, & Welser, 2006; Turner, Smith, Fisher, & Welser, 2005), which helped to illustrate the benefits of the approach. While network data is important, it provides just one perspective, which can be supported with the addition of other visualisations. As suggested by Fernanda Viégas, "we usually understand things better we have more than one way of looking at them" (F.B. Viégas & Donath,

2004).

3 DOMAIN ANALYSIS

Initially, we presented some of our previous work to the team in Symantec. We described the aim of this work, which was to provide reflective support to online communities (McAuley, O'Connor, & Lewis, 2012), and discussed, in broader terms, how we thought visualisation could support the role of the community manager. In a way, much of this discussion was driven from the literature, 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 of community management had not been formalised into strategic or operational directives. Also, in some cases, the notion of visualising a community to support the management process 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 process as it was to elicit a set of design requirements.

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 principle researcher, and were aimed at understanding the team's management processes and their need for a visual analytics tool. The third and final workshop included an additional member of the team and a second researcher, and was used to present and get feedback on initial designs, and to identify a feature set for clustering users in the community into a set of cultures. While we aimed to video each workshop, some of the team felt more comfortable with audio, thus the audio was recorded for each workshop. In this section, we present and discuss the findings from each of the three workshops.

3.1 First Workshop

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

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

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

3.1.1 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, 2000), but to support and encourage pro-social behaviour. The aim, it was 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, she has only banned one user. Rarely, have 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. A "promoter", on the other hand, is someone who either promotes the company's products and services or helps out in the forum.

Aside from converting detractors to promoters, members of the team were 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. Thus peers, regular or casual community members, provide the required technical support, 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 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. 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, the online community strategy.

3.1.2 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 participants do not find this measure particularly useful however. 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 posts to number of threads, 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 one thread posting negatively about a specific topic.

3.1.3 How Do You Envision Visual Analytics Supporting the Community Manager?

In answering this question, the team identified four main objectives for the visual analytics application:

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. The difficulties of sentiment were discussed at some length. A member of the team, who had experience of sentiment analysis in other projects, suggested avoiding this approach entirely due to the "inaccuracies" and difficulties that it presented previously. They suggested that the information provided in the visualisation 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 (Gazan, 2009; Suh, Chi, Pendleton, & Kittur, 2007), was suggested by the team.

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 in terms of answers, to the community. One of the team mentioned that while she knows her community very well, she felt that there could still be users, who have untapped potential, and thus slip through the net because she was unable to identify that potential early enough. Being able to identify users that could provide a greater contribution to the community is important.

Cultures

The team were interested in partitioning the community into several different "cultures" with each culture representing a degree or level of contribution to the community. To take a simple example from the first workshop, as illustrated in Figure 1, 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 addressed in the literature on the identification of social roles (Chan & Hayes, 2010; Crowston & Howison, 2005; Golder & Donath, 2004; Kelly, Fisher, & Smith, 2006; Malouf & Mullen, 2008; Welser et al., 2011; Welser, Gleave, Fisher, & Smith, 2008). The difference with this approach is that the team was interested in identifying users that have the capacity to contribute to the community in a more meaningful way and then converting those users from one culture to the next over a sustained period of time. In this capacity, 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. While similar frameworks have been proposed, most notably Preece and Shneiderman's reader-to-leader framework (Preece & Shneiderman, 2009), we have yet to find an actual implementation of this approach in the online community literature.

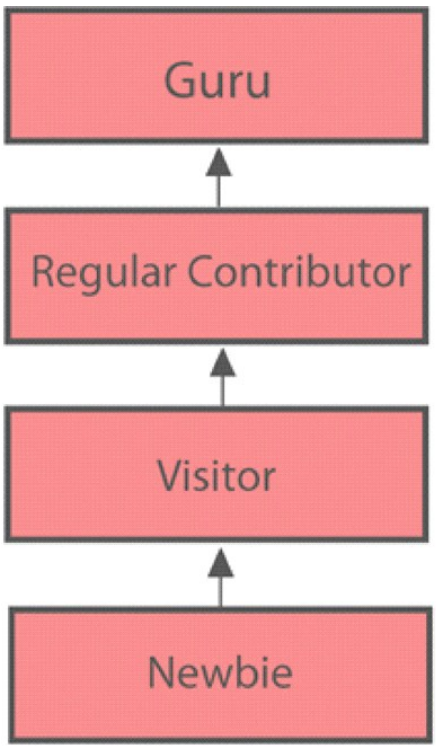

Figure 1: An illustration of community cultures, as presented by one of the team in the first workshop In this image, a user starts out as a newbie Over time, that user becomes a visitor, develops into a regular contributor and then emerges as a guru This is an "ideal" trajectory, from the perspective of an online community manager, for a user in a technical support forum

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, she could use the path visualisation to confirm and thus validate her 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 the online community strategy.

In the second stage of the workshop, the team sketched some ideas on a whiteboard that were then discussed as a group (see Figure 2). 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 tool. From there the community manager could select users that are of interest and then "carry her analysis forward" by comparing user paths using other visualisations methods. Coordinating these visualisations (North, 2000) was suggested but discarded, due to screen dimensions and the quantity of information being presented.

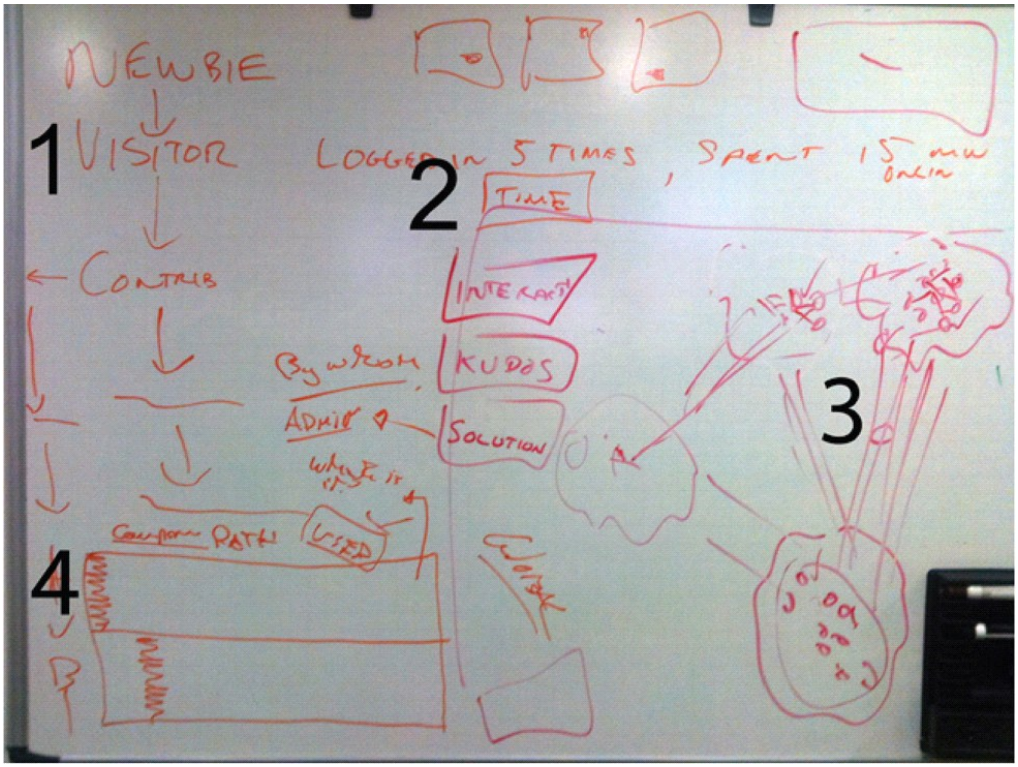

Figure 2: The results from the second stage of the first workshop held with the community management team General ideas and initial designs as proposed during brainstorming 1. Cultures as articulated by the participants in the first workshop Cultures are defined by their levels of contribution, from 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

3.1.4 Implications

Certainly, the outcome of this session has several implications for the role of the community manager and the practice of online community management: 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 intervention 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 intervention with the community? Are there gold standard approaches to encouraging certain users or dealing with specific cultures? Thirdly: Visual analytics will be considered a key component of the community management process, enabling the community manager to not only analyse the activity of the community but also assess their own interventions, and the result of those interventions, with the community.

4 DESIGN REQUIREMENTS

Having completed the first workshop, we drafted a set of design requirements.

- **R1 Multiple 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 she explores the global visualisation to select users of interest and then confirms her selection using the local visualisation. Over time, the global visualisation can support both kinds of tasks, the global view confirming that a user has migrated into a new culture while also allowing the community manager explore the community. In this way, the manager can build a more comprehensive mental-model of the how the community is organised and how specific users contribute to the community.
- **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. While there is a lot of work being conducted on identifying social roles in online communities, the aim of this approach is to partition users by their level contribution coupled with the level of peer recognition. This is a linear trajectory that extends from users who are barley active to the highest and most well respected contributors in the community.
- **R3 Paths**: The team were interested in visualising a user's path, their journey or trajectory, through the community over time. This is "a logical path, which is based on increased levels of activity."
- **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.
- **R5 Outliers**: As with the use of most 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 but considered superfluous to the analytic process. Parameter fine-tuning, whereby a user adjusts the feature set or the parameters of the clustering algorithm, was discussed but considered as excessively complex, considering the domain analysis, and would require further investigation beyond the scope of the current study. However, the ability to pan, zoom, search and filter was requested to enable traversal of the dataset.
- **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, and so forth.
- **R8 Selection:** The team requested the ability to select users that could be then "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.
- **R10 Design**: Finally, the team suggested that the application should have a certain design aesthetic. This was considered important if the tool was to be used as part of a broader 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 possibility of 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.

5 DESIGN WORKSHOPS

In the next two workshops, we focused on producing an initial set of designs, validating those designs as mock-ups presented to the team, selecting a set of features for the clustering process and then validating the cultures from the perspective of the community manager.

5.1 Second Workshop

In the second workshop, the researcher 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. Only a couple of the examples were related specifically to online communities. Having completed the presentation, the researcher handed out an information sheet 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 and a textual description of the visualisation on the right as illustrated in Figure 3.

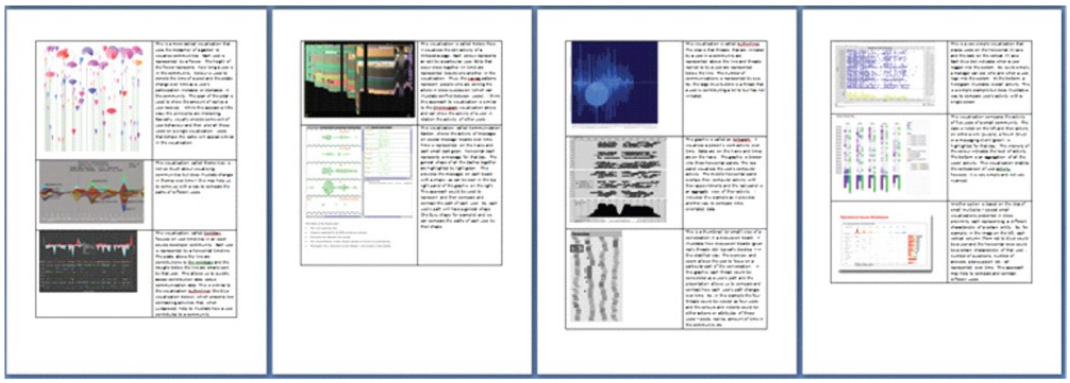

Figure 3: A sheet of visualisation examples used to guide discussion and present alternative approaches during the second workshop

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 alternative visualisations.

5.2 Third Workshop

The third workshop had two specific goals: To validate our initial designs (as illustrated in Figure 4) 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. Very obscure measure renders the visualisation less comprehensible, thus it was important to create a model that community managers can understand. Following the workshop, we drafted up all the metrics, numbering over 60 in total, which could be used as features in clustering. This was then reduced to a smaller set as discussed in more detail in section 6.1.1.

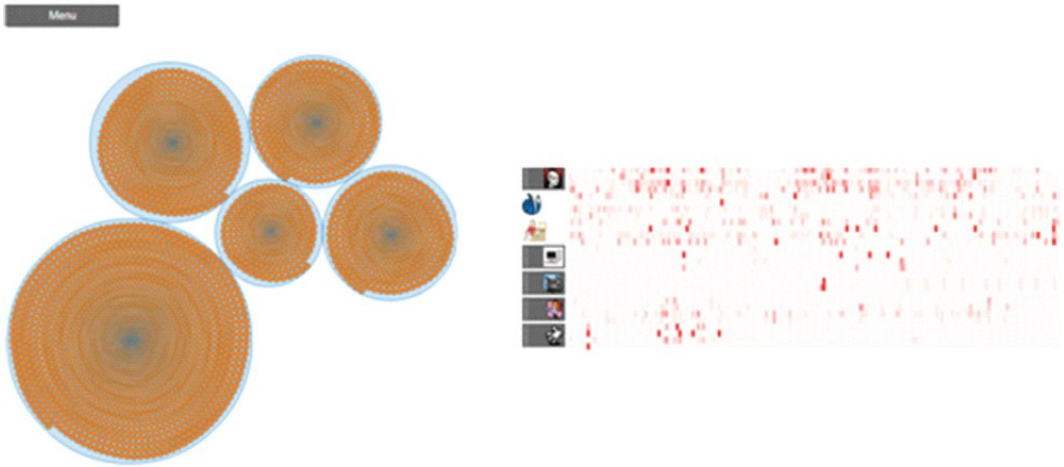

Figure 4: Some initial mock-ups The cultures visualisation (on the left) is close to how the cultures are represented in later versions of Petri The paths visualisation (on the right) was rejected as noisy and non-intuitive In this visualisation, the users are represented by profile images on the left and their activity is presented like a heat map stretching left to right

As regards the visualisations, the team felt comfortable with how the cultures were represented (see Figure 4). 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. 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 4). 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. It was described as non-intuitive and overly complex. It was suggested that a more familiar approach be considered. As result, we discussed timelines, sparklines and small multiples as an alternative and more familiar representation.

6 DESIGN OF PETRI

Petri is an initial attempt to develop a visual analytic application that can support, and thus be integrated into, the work practice of an online community manager. In this section, we present the application's main features. We describe each visualisation, justifying our design decisions with reference to the user requirements (see section 4). We also present our approach to partitioning and then ordering the community into number of discrete cultures.

6.1 Visualsiation1: Cultures

Initially, we choose a scatter plot, varying the hue of visual variable according to their culture (Bertin, 1983). 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, at the point of "inventory and classification" (Bertin, 1983b) for example, 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 model of the community, particularly as the tool is to be used at regular intervals over time.

Comparing change (R9) in cultures over time is not addressed in this chapter, but, as discussed, this requirement contributed to the choice of technique. The culture 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. This is not reported in this chapter however; as we did not address it in the evaluation, and is thus considered in future work.

We re-organised the layout and composition of the visualisation. We drew together each culture, as illustrated in Figure 5 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 discussed 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 click on, 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. How useful this information is to a community manager, who is not an experienced data analyst, is speculative, and could serve

to confuse rather than clarify their analysis.

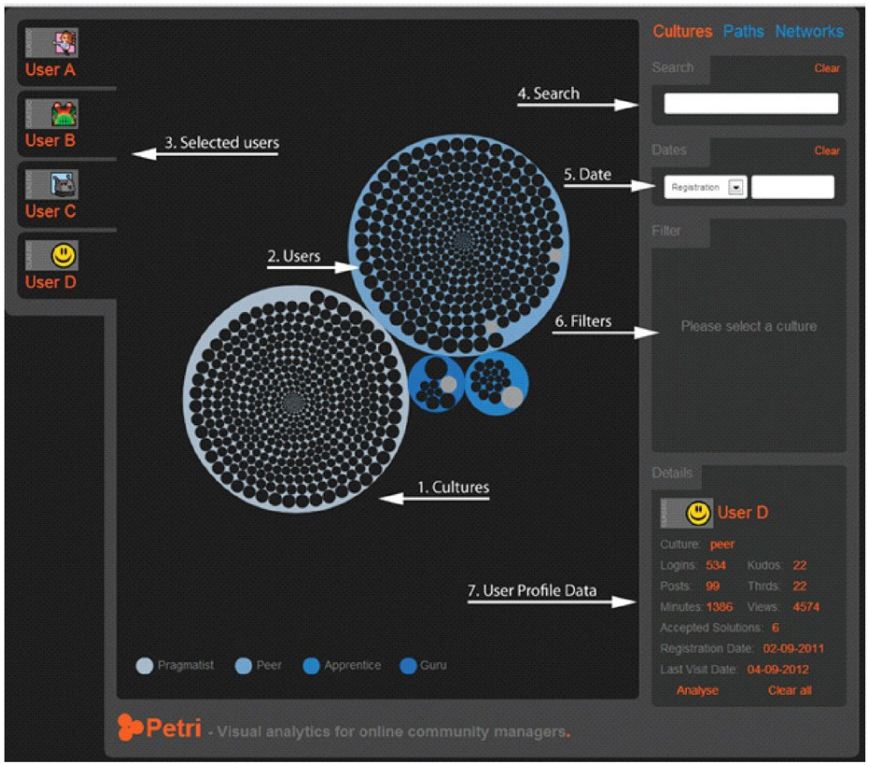

Figure 5: Petri Cultures Visualisation 1. Interactive dispaly with different cultures organised in a clockwork configuration: pragmatist, peer, apprentice and guru, 2. Individual users, organised according to the distance from the medoid of their cluster The larger the glyph, the further the user is 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 communty manager zooms into a culture), 7. User Profile Data (R7)

6.1.1 Algorithm

The aim of the cultures visualisation is to partition the community into different cultures and visualise those cultures for the online community manager. Although the term "culture" was proposed by participants in the first workshop, the term social role is more commonly used in the online community literature. Generally, social roles are informal social positions taken up by users in an online community and are 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). However, over the last number of years, researchers have used a combination activity and social network visualisations to reveal a number of additional social roles, such as answer person, debater (Usenet) and technical editor (Wikipedia) (Fisher et al., 2006; Kelly et al., 2006; Viégas & Smith, 2004; Welser et al., 2011; Welser et al., 2008). Our conception of a culture may be a little different from the social role proposed by Golder and Donath in that we are interested in identifying points along a linear trajectory (see Figure 1) that can be used to gauge a user's increasing level of contribution.

Initially, we followed the approach proposed by Chan and Hayes (Chan & Hayes, 2010) 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, we found it difficult to communicate the nature of these user roles to the community management team because they are based on a combination of complex (social network) measures. Also, there is no indication of a clear trajectory, which users can follow from lurker to guru using this approach as the categorisation is skewed by the structural position of the user in the community. Consequently, we removed the social network measures, and concentrated, instead, on identifying a number of activity based ratios that could be used to illustrate an increasing level 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 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 (fourth session). 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 the supplementary features can be used to replace some of the initial set.

The resulting 15 features are shown in Table 1.

Table 1: 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	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, our aim was to remove marginal users form the population, thus "cutting the tail", and then partitioning the remaining distribution into a number of discrete yet ordered "bands", with each band representing an increasing level of user 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 converted all the above features into percentages of the overall population and then normalised the dataset. So, for example, the percentage replies of a specific user was calculated by comparing the number of replies that 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. 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 (see Table 2). 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 2: 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", 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 1. 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.

This resulted in 4 cultures:

1. **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.

2. **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.
3. **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.
4. **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.

6.1.2 Interaction

We sought to provide straightforward and intuitive ways for the user to improve the readability and representation of the visualisation (R6). Thus, the user can pan and zoom, 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 6). 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, who is 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 the community in a intuitive manner. We felt that this was important. 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 refinement is required before implementation and evaluation of this sort of interaction.

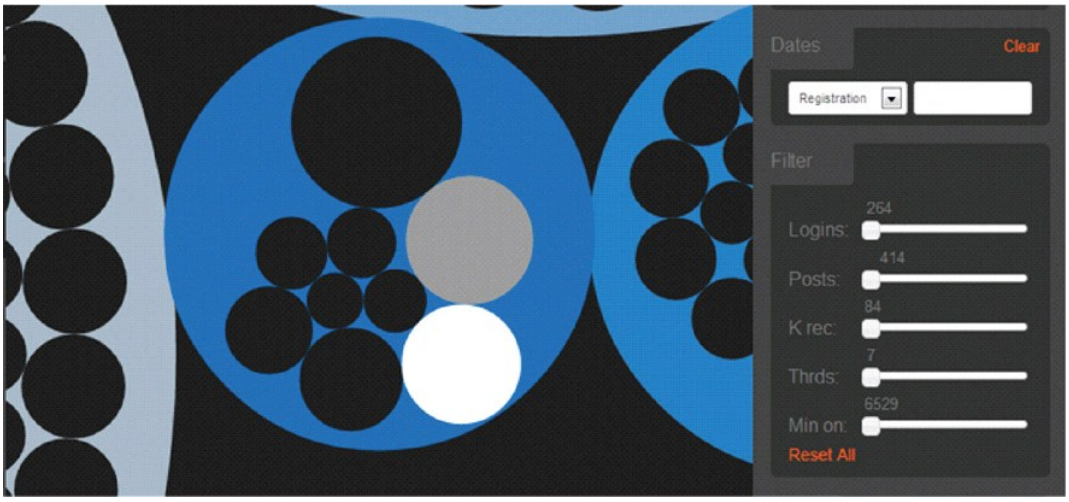

Figure 6: Cultures visualisation zoomed in on guru culture users filtered on login

6.1.3 Selection

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 (see Figure 7). Selected users are presented to the community manager in the top left of the visualisation interface and are persisted across the cultures, paths and network visualisations. At both the cultures and networks interface, the community manager is able to add additional users for analysis.

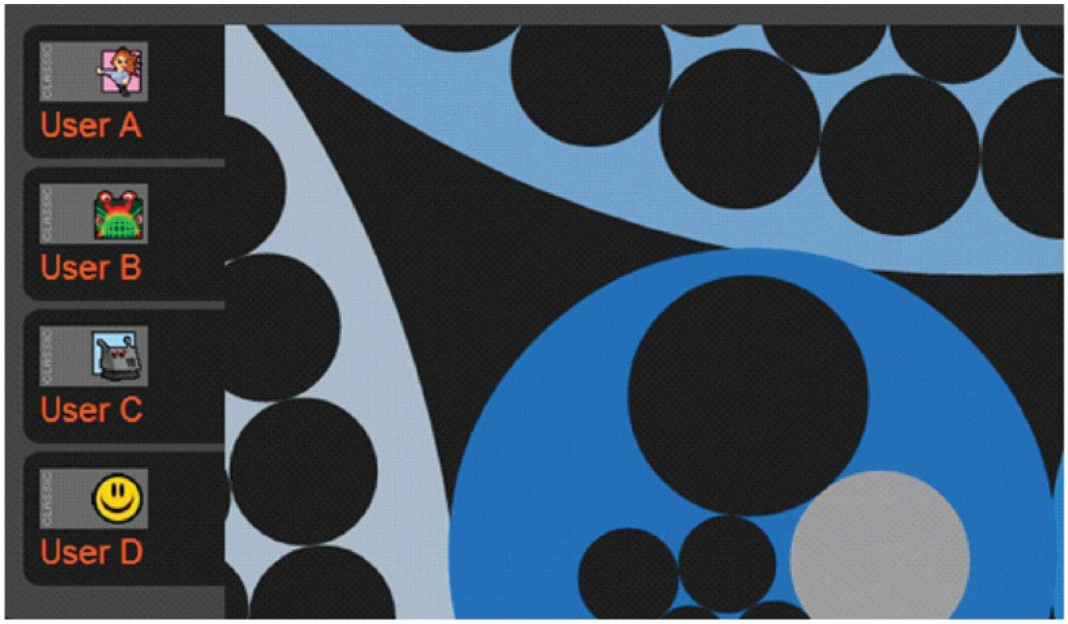

Figure 7: Selection (R8)-Users selected for analysis by the community manager

6.2 Visualisation 2: Paths

In Petri, the community manager is able to analyse and compare how users develop 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, 1983c). 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 forum began). We avoided cumulative frequencies so that the graph would present the general tendency of a user's contribution since joining the community (see Figure 8).

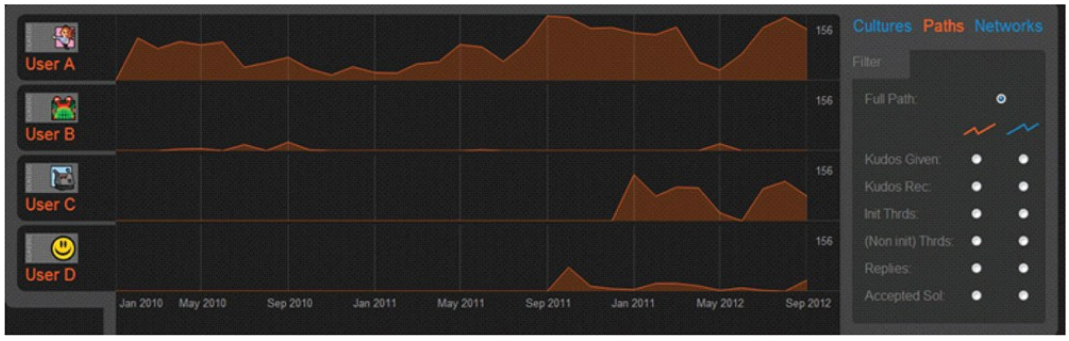

Figure 8: Path visualization. In this image, the communtiy manager is comparing the combined contirbution rates of four users The user's path (considered as an aggregation of all productive activtiy in the community) is presented on a time-series At a glance the most active contributors are perceptable, however, the approach also allows the community manager compare their contribution rates with other users

6.2.1 Interaction

While this approach facilitates comparison, and supports confirmatory analysis, 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 section 2.1, and introduced a third component (see Figure 9) as an inverted time-series positioned under the x-axis. Now the visualisation resembles a stacked graph (Heer, Bostock, & Ogievetsky, 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. We used animation to transition between the time-series (Complete Paths) and the stacked graph

(Individual Actions).

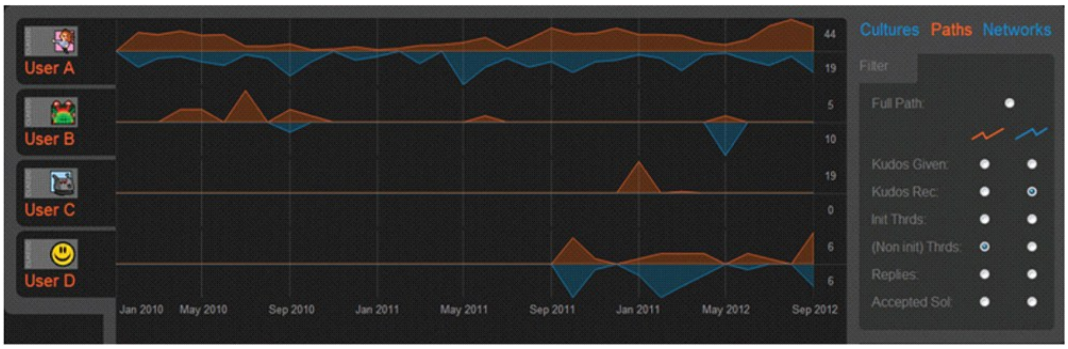

Figure 9: Filtered path visualisation In this image, the community manager is comparing the paths of four different users Non-initialised threads are in orange and received kudos is in blue The representation enables the community manager to quickly evaluate how useful a user's contribution is considered by the community over time

6.3 Visualsiation 3: Ego-Centric Networks

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 (graphs of over 150 nodes can be problematic). 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.

6.3.1 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 10). 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. In this network visualisation, the community manager can also select other users from the network and examine then their paths and networks (R8).

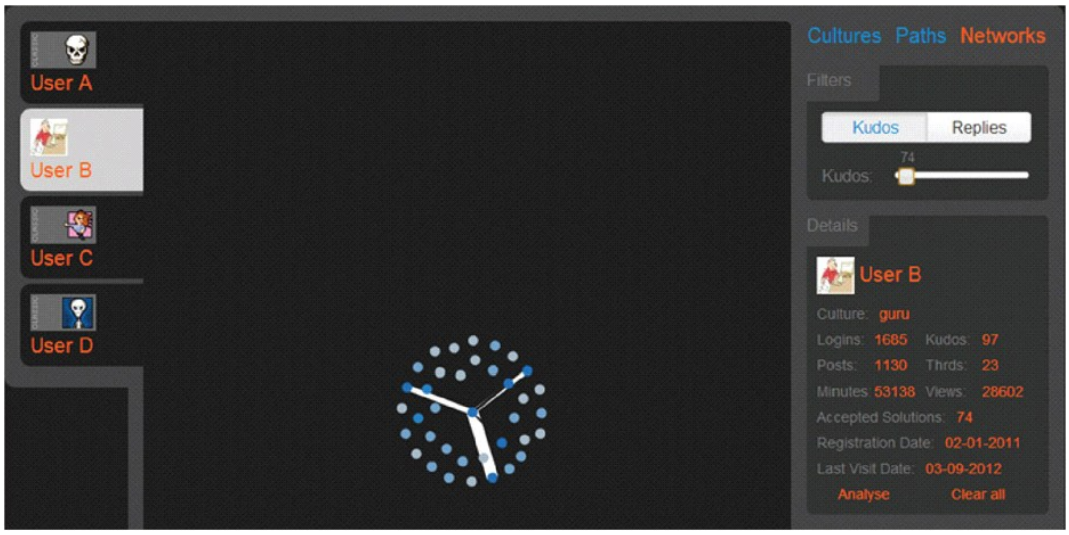

Figure 10: Network Visualisation In this image the user is analysing the ego-centric networks of four users She is looking at their kudos network and has filtered out any connections that are fewer than 74 kudos

7 EVALUATION

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 design sessions. 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. The aim of the

evaluation was to critically asses the usability and usefulness of the Petri visual analytics tool.

7.1 Methodology

We drew on three sources to help define our methodology. The first was Ellis and Dix's approach to exploratory analysis in which they argue for less performance-based, in favour of more exploratory-based, evaluation strategies for information visualisation systems. In this paper, domain experts are considered to as "worth their weight in gold" (Ellis & Dix, 2006) and exploring the visualisation without predetermined tasks is favoured. The second was Isenberg et al.'s paper on grounded evaluation strategies, in which the authors argue for conducting the evaluation in the context of its intended use (Isenberg et al., 2008). Both papers suggest a qualitative, and chiefly observational, approach to evaluation, which is germane giving the size and required expertise of the user group. The third and final source is Andrew's short paper on evaluating information visualisation (Andrews, 2006) in which he suggests selecting the "correct users" and applying the "correct tasks" on a stable visualisation environment.

7.2 Procedure

The five sessions were conducted back-to-back and each session was videoed for analysis later. First, we asked three preliminary questions to establish the participant's familiarity with the community and, more generally, their understanding of visualisation tools 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 requires intervention from a facilitator. 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. 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, North, & Duca, 2005; Yi, Kang, Stasko, & Jacko, 2008) they come across 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 table in (Benedek & Miner, 2002). 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 an approach similar to (Heer, Viégas, & Wattenberg, 2009), to first code and then evaluate the usefulness 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 use North's insight based methodology, 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.

7.3 Results

In this section, we present the results of the video analysis, and then we discuss the post evaluation interview. We were interested in understanding how useful and usable participants found Petri but also aimed to assess whether any observational cycles were evident in how they went about their analysis. Finally, we identified, and categorised, any usability issues that participants encountered during their session.

7.3.1 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, primarily, to find users that she found interesting. Having participated in earlier design sessions, 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 judiciously. 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? Can Petri help a community manager make decisions about their community? What actions would a user suggest having used Petri? We triangulated answers from the video analysis and exit interview.

The large majority of comments that participants made during the evaluation were observational (Figure 11), 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 she "wanted to see how he was doing". She had contact with him before and thought he might be "Guru material". She found the user and then analysed his path. She 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. She 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." This helps to validate her decision as regards the potential of this user. She 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 validate the selection process qualitatively, through interaction with users, and quantitatively, through analysis with Petri, helps to act against these sorts of problems.

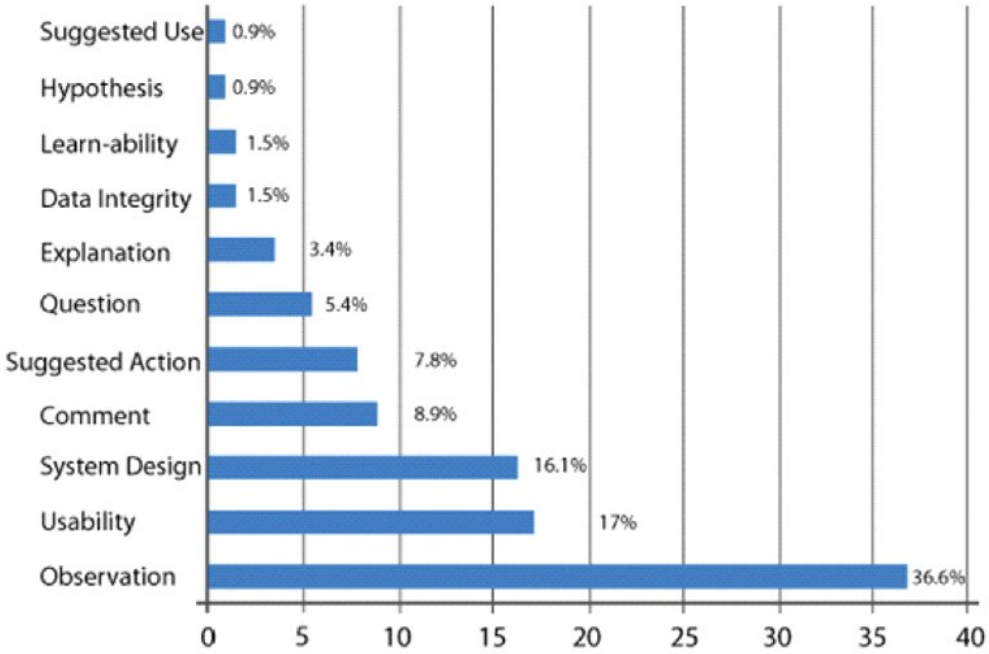

Figure 11: Content analysis of transcribed video Categories are not mutually exclusive

Similarly another participant filtered for users who just joined the community this year. Finding a user that is particularly active, she says "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 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.

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. She said she preferred the time-series as this was an approach she 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. 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%).

7.3.2 Interviews

Although the analysis helped us evaluate the utility of Petri, we also wanted to better understand the context in which the tool would be applied. We conducted a short exit interview with each participant, in which we drew upon the themes of utility, design, improvement and the process of community management. Here we briefly discuss the interviews:

Usefulness

All users indicated that they found Petri to be useful. Several focused on explicit aspects of Petri. 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 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 her much more "visibility" into the community and would provide grounds for validating the decision-making of her community mangers.

Design and Presentation

Not many participants used the network visualisation. One participant said that she found ego-centric network too restrictive, she said "it might

be too focused on the individual, it does not tell me anything about whole network look like", and that she 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 mentioned.

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 Petri, 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. She suggested that, at present, it looks as if these three functions 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. The concept of "linking" was proposed by several participants, whereby the user could easily traverse between Petri and the forum whilst maintaining the context of analysis. This is important for a community manger to assess the "quality" of a user's contribution, and while similar approaches have been implemented in previous work (Zhu & Chen, 2008), this feature was not discussed in previous sessions. 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.

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.

The Process of Community Management

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 in practice. Reflecting on this, one of the participants 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.

Another 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. She 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 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 she would use the tool in practice:

> 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.

She 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. She suggested that adoption of such advanced visual analytic tools could meet with some resistance from those less technical.

8 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, abstraction, interaction and algorithmic design. While our aim was to pursue this model 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 encodings) and to identify alternative, and possibly more

adept, interactive strategies (addressing interaction). 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 helped formalise our design process further. For the reader, a taxonomic approach could help clarify certain design decisions, while for the reviewer; such an approach could provide the grounds for better critical analysis. In future work, this is something that the authors will address. Our approach to algorithmic design was initially presented in the third workshop, as discussed, and then later refined through discussions (both face-to-face and over email) with 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 entire design. The evaluation conducted at the end of the process helped to assess the degree of usefulness of the overall system.

There are also some limitations as regards to the generalisability of Petri. While the community in question is small, given the scope of the organisations other online communities, participants had asked in early workshops could this tool be easily ported to other communities? Certainly, the clustering 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. 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. We developed Petri to dovetail into the community management process, so that the community manager could regularly asses the results of their intervention with the community. Such an approach may be much less possible with large online communities; however, this could be addressed with future work.

9 FUTURE RESEARCH DIRECTIONS

Immediate directions for future work include the implementation of small multiples (Tufte, 2001), that will support the analysis of change in the community's cultures over time. It was our intention to pursue this aspect of the application having completed an initial evaluation, as it can be evaluated independently of the rest of the application. Also, integration of Petri with the actual forum software, to allow the community manager to qualitatively substantiate any quantitative assessment is also required. This has been raised in early studies (Viégas & Smith, 2004) but was not addressed in this implementation. Furthermore, given the negative reception to the ego-centric visualisation there is scope to re-address how to convey network information in Petri. Certainly, approaches such as Vizster, discussed in section 2.2, could be useful in this context.

While we have partitioned the community in quite a coarse grained manner, there is scope to develop a categorisation scheme that operates at a much finer level of resolution. This could be a significant contribution to the analysis of online communities; as such an approach can be reapplied to any forum that has a similar socio-technical infrastructure–users, boards, threads, replies and kudos. In fact, this data structure makes up for the vast majority of question and answer communities on the web. Furthermore, we did not provide the ability for the community manager to alter the parameters of the clustering algorithm used to partition the community. Advancing analytics in this direction, so that non-professional analysts can progressively carry-out increasingly complex computational tasks, within, of course, a clearly defined scope, is not well understood at present. However, as visual analytics shifts into the main stream, this presents an opportunity for future work.

The ultimate indication of success for Petri is adoption and it is clear from the evaluation that more work is required before Petri could be adopted into the workflow of the community manager. It is also clear, however, that given the busy schedule of a professional community manager, it is difficult to assess how visual analytics could be successfully adopted in practice. While participants recognised the benefits of Petri, and some even suggested re-addressing the process of online community management given the application of such tools, the need to provide solutions that improve productivity is always pressing. Moreover, to understand if visual analytic tools, such as Petri, can support the role of community manager on a long term basis, and provide a more analytic approach to online community management, further, principally longitudinal, evaluation is required.

Finally, as regards the methodology, while Muzner's model supports an informed design process-one that is both aware of the design space and the incremental procedures required to construct a useful visual analytics application - the model does not account for deployment or address how the tool is adapted when used in practice. In most cases there is a need to amend the application, or at least alter varying aspects of the application, once deployed. While a focus on design is precedent, it is also important to factor in the social and organisational issues that can impact the success of any deployment. Other models, such as (Dou, Butkiewicz, Bier, & Ribarsky, 2011), may prove a useful addition in this context.

10 CONCLUSION

In this chapter we described a case study in which we designed, developed and evaluated a visual analytics tool to support the management of online communities. The tool was designed to encourage a more analytic approach to online community management, which is based on cycles of observation and intervention. In designing the tool we adopted Muzner's nested methodology for visualisation design and validation and conducted several interviews and design workshops with Symantec's online community team. From these sessions we drafted a set of requirements that were then used to inform the design of the resulting visual analytics application, called Petri. Petri enables the community manager to analyse their community from multiple perspectives, shifting between phases of explorative and confirmative analysis, and to identify users that could prove valuable to the community as it evolves over time. We presented the results of an explorative evaluation, conducted with five members from the community management team, and found the visualisation tool was both useful and usable. Nevertheless, questions were raised in exit interviews as to how the tool would perform in practice.

We made the following contributions:

- First, we proposed a novel approach to online community management, which is based on cycles of observation and intervention, and is supported by the application of advanced visual analytic tools.
- Second, we presented a set of design requirements that can be readdressed by other researchers interested in online community visualisation.
- Third, we presented Petri, a novel visual analytics tool that was designed, developed and evaluated with Symantec's online community management team.

Finally, we outlined several directions for future work. While the evaluation presented in the paper helps to establish the utility or the tool, there is scope to evaluate how this application would perform in practice. Adoption is not based solely on utility and usability, however, but is also based on the social and organisational context in which the tool is deployed. To shift visual analytics from the research community into the hands of community practitioners there is a need to address not only how the tool is designed but also how the tool is deployed in practice. This, of course, requires a greater degree of commitment from an online community manager and would include a longitudinal research methodology such as Multi-dimensional In-depth Long-term Case Study (MILC), as outlined in Shneiderman and Plaisant (Ben Shneiderman & Plaisant, 2006).

ACKNOWLEDGEMENT

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Trinity College Dublin. Thanks to members of the Symantec community management team for their participation in this research.

REFERENCES

Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. New York: Hyperion.

Andrews, K. (2006). Evaluating information visualisations. In *Proceedings of the 2006 AVI Workshop On Beyond Time And Errors: Novel Evaluation Methods For Information Visualization*, 1–5. New York: ACM Press.

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. doi:10.1109/TVCG.2010.78

Bacon, J. (2009). *The art of community: Building the new age of participation (Theory in practice)*. Sebastopol, CA: O'Reilly Media.

Bederson, B. (1999). Does animation help users build mental maps of spatial information? In *Proceedings of the 1999 IEEE Symposium on Information Visualization, 11*, 28. Washington, DC: IEEE Press.

Benedek, J., & Miner, T. (2002). Measuring desirability: New methods for evaluating desirability in a usability lab setting . In *Proceedings of Usability Professionals*. New York: ACM Press.

Bertin, J. (1983a). *The retinal variables. Semiology of Graphics: Diagrams, Networks, Maps (65)*. Madison, WI: The University of Wisconsin Press.

Bertin, J. (1983b). *Colour Variation. Semiology of Graphics: Diagrams, Networks, Maps (91)*. Madison, WI: The University of Wisconsin Press.

Bertin, J. (1983c). *Diagrams involving two components. Semiology of Graphics: Diagrams, Networks, Maps (195)*. Madison, WI: The University of Wisconsin Press.

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 *Proceedings of International Conference on Computational Science and Engineering*, 332–339. New York: ACM Press.

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*(3.2), 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*, 1060–1069. Washington, DC: IEEE 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., Sproull, L., Kiesler, S., & Kraut, R. (2007). Community effort in online groups: Who does the work and why . In S. PWeisband, (Ed.), *Leadership at a Distance* (171–195). Florence, KY: Psychology Press.

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*, 92–99. Washington, DC: IEEE Press.

Chan, J., & Hayes, C. (2010). Decomposing discussion forums using user roles. In *Proceedings of the Second Web Science Conference*. Raleigh, NC: ACM Press.

Crowston, K., & Howison, J. (2005). The social structure of free and open source software development. *First Monday*, *10*(2). doi:10.5210/fm.v10i2.1207

Donath, J. (2007). Signals in social supernets . *Journal of Computer Mediated Communication*, *13*(1).

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

Dou, X. W. W., Butkiewicz, T., Bier, E. A., & Ribarsky, W. (2011). A two-stage framework for designing visual analytics system in organizational environments. In *Proceedings of 2011 IEEE Conference on Visual Analytics, Science, and Technology,* 251–260. Providence, RI: IEEE Press.

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*, 109–116. Venice, Italy: AVI Press.

Elmqvist, N., Do, T.-N., Goodell, H., Henry, N., & Fekete, J.-D. (2008). ZAME: Interactive Large-Scale Graph Visualization. In *Proceedings of IEEE Pacific Visualization Symposium,* 215–222. Kyoto: IEEE Press.

Erickson, T. (2003). Designing visualizations of social activity: Six claims. In *Proceedings of CHI EA '03 CHI '03 Extended Abstracts on Human Factors in Computing Systems*, 846–847. Ft. Lauderdale, FL: ACM Press.

Fisher, D. (2005). Using egocentric networks to understand communication. *IEEE Internet Computing*, *9*(5), 20–28. doi:10.1109/MIC.2005.114

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.* Washington, DC: IEEE Press.

Gazan, R. (2009). When online communities become self-aware. In *Proceedings of Conference on Hawaii International System Sciences.* Washington, DC: IEEE Press.

Gilbert, E., & Karahalios, K. (2007). CodeSaw: A social visualization of distributed software development. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction-Volume Part II*, 303–316. New York: HCIR Press.

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.

Golder, S. A., & Donath, J. (2004). Social roles in electronic communities. *Internet Research*, *5*, 19–22.

Gruzd, A., & Haythornthwaite, C. (2008). The analysis of online communities using interactive content-based social networks. *In Proceedings of the American Society for Information Science and Technology Conference*, 523–527. Paris: ICSU Press.

Heer, J., 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; Tomorrow's Computing Today*, *8*(5), 20.

Heer, J., & Boyd, D. (2005). Vizster: Visualizing online social networks. In *Proceedings of IEEE Symposium on Information Visualization,* 32–39. Minneapolis, MN: IEEE Press.

Heer, J., & Hellerstein, J. (2009). Data Visualization and social data analysis. In *Proceedings of the VLDB Endowment,* 1656–1657.

Heer, J., & Perer, A. (2011). Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *Proceedings of 2011 IEEE Conference on Visual Analytics Science and Technology*, 51–60. Providence, RI: IEEE Press.

Heer, J., Viégas, F., & Wattenberg, M. (2009). Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM-Rural Engineering Development*, *52*(1), 87–97. doi:10.1145/1435417.1435439

Henry, N., & Fekete, J.-D. (2006). MatrixExplorer: A dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 677–684. doi:10.1109/TVCG.2006.160

Holten, D. (2006). Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 741–748. doi:10.1109/TVCG.2006.147

Husson, F., Le, S., & Pages, J. (2010). *Exploratory multivariate analysis by example using r*. New York: CRC Press. doi:10.1201/b10345

Isenberg, P., Zuk, T., & Collins, C. (2008). Grounded evaluation of information visualizations. In *Proceedings of the 2008 Workshop on Beyond time and errors: Novel Evaluation Methods for Information Visualization*, 1–8. Florence, Italy: ACM Press.

Kelly, J. W., Fisher, D., & Smith, M. (2006). Friends, foes, and fringe: Norms and structure in political discussion networks. In *Proceedings of the 2006 International Conference on Digital Government Research*, 412–417. San Diego, CA: ACM Press.

Kimmerle, J., Moskaliuk, J., Harrer, A., & Cress, U. (2010). Visualizing co-evolution of individual and collective knowledge. *Information Communication and Society*, *13*(8). doi:10.1080/13691180903521547

Kittur, A., & Kraut, R. E. (2010). Beyond wikipedia: Coordination and conflict in online production groups. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, 215–224. Savannah, GA: ACM Press.

Lohr, S. (2012). The age of big data. *The New York Times*. New York.

Malouf, R., & Mullen, T. (2008). Taking sides: User classification for informal online political discourse. *Internet Research*, *18*(2), 177–190. doi:10.1108/10662240810862239

Marin, A., & Wellman, B. (2010). Social network analysis: An introduction . In Carrington, P., & Scott, J. (Eds.), *Handbook of Social Network Analysis*. London: Sage.

Masli, M., Priedhorsky, R., & Terveen, L. (2011). Task specialization in social production communities: The case of geographic volunteer work. In *Proceedings of ICWSM*, 217–224. Palo Alto, CA: AAAI Press.

McAuley, J., O'Connor, A., & Lewis, D. (2012). Exploring reflection in online communities. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 102–110. Vancouver, BC: ACM Press.

Morzy, M. (2009). On mining and social role discovery in internet forums. In *Proceedings of 2009 International Workshop on Social Informatics*, 74–79. Warsaw, Poland: IEEE Press.

Munzner, T. (2000). *Interactive Visualization of Large Graphs and Networks.* (Unpublished doctoral thesis). Stanford, CA, Stanford University.

Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 921–928. doi:10.1109/TVCG.2009.111

North, C. (2000). Snap-together visualization: Can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies*, *53*(5), 715–739. doi:10.1006/ijhc.2000.0418

North, C. (2006). Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, *26*(3), 6–9. doi:10.1109/MCG.2006.70

O'Keffe, P. (2008). *Managing online forums: Everything You need to know to create and run successful community discussion boards, 320*. Saranac, NY: AMACOM.

Panciera, K., & Halfaker, A. (2009). Wikipedians are born, not made: A study of power editors on Wikipedia. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, 51–60. Sanibel Island, FL: ACM Press.

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

Preece, D. J. (2000). *Online communities: Designing usability and supporting sociability*. Hoboken, NJ: Wiley.

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.

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), 1325–1332. doi:10.1109/TVCG.2008.125

Saraiya, P., North, C., & Duca, K. (2005). An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *11*(4), 443–456. doi:10.1109/TVCG.2005.53

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

Shneiderman, B., & Aris, A. (2006). Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 733–740. doi:10.1109/TVCG.2006.166

Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 1–7. Philadelphia: ACM.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., & Capone, T. (2009). Analyzing (social media) networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies-C&T '09*, 255. New York: ACM Press.

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*, 1037–1040. Florence, Italy: ACM Press.

Suh, B., Chi, E. H., Pendleton, B. A., & Kittur, A. (2007). Us vs. them: Understanding social dynamics in wikipedia with revert graph visualizations. In *Proceedings of 2007 IEEE Symposium on Visual Analytics Science and Technology*, 163–170. Washington, DC:

IEEE Press.

Tory, M., & Moller, T. (2004). Rethinking visualization: A high-level taxonomy. In *Proceedings of IEEE Symposium on Information Visualization*, 151–158. Washington, DC: IEEE Press.

Tufte, E. R. (2001). *Data denstiy and small multiples. The Visual Display of Quantitative Information (42)*. Cheshire, CT: Graphics Pr.

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing usenet: Mapping computer-mediated collective action . *Journal of Computer Mediated Communication*, *10*(4).

Tversky, B., Bauer, J., Betrancourt, M., De Europe, A., & St-Martin, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247–262. doi:10.1006/ijhc.2002.1017

Van Ham, F., Schulz, H., & Dimicco, J. M. (2009). Honeycomb: Visual analysis of large scale social networks. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II*,429–442. Uppsala, Sweden: Springer.

Vassileva, J., & Sun, L. (2007). Using community visualization to stimulate participation in online communities. *E-Service Journal*, *6*(1), 3–39. doi:10.2979/ESJ.2007.6.1.3

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 *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 109–118. Washington, DC: IEEE Press.

Viégas, F. B., & Donath, J. (2004). Social network visualization: Can we go beyond the graph. In *Proceedings of Workshop on Social Networks, 4*, 6–10. Chicago: ACM Press.

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.* Washington, DC: IEEE Press.

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*, 575–582. Vienna, Austria: ACM Press.

Viégas, F. B., Wattenberg, M., Kriss, J., & Ham, F. V. (2007). Talk before you type: Coordination in wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* (78). Hawaii: IEEE Press.

Wagner, C., Rowe, M., Strohmaier, M., & Alani, H. (2012). Ignorance isn't bliss: An empirical analysis of attention patterns in online communities. In *Proceedings of IEEE International Conference on Social Computing.* Amsterdam: IEEE Press.

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*, 517–520. New York: ACM Press.

Wattenberg, M., & Kriss, J. (2006). Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, *12* (4), 549–557. doi:10.1109/TVCG.2006.65

Wattenberg, M., Viégas, F., & Hollenbach, K. (2007). Visualizing activity on wikipedia with chromograms. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction-Volume Part II*, 272–287. Rio de Janeiro: Springer.

Welser, H., Cosley, D., Kossinets, G., & Lin, A. (2011). Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, 122–129. Seattle, WA: JALT.

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*.

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,* 37–44. Asheville, NC: ACM Press.

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 *Proceedings of the 2008 Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 1–6. Florence, Italy: ACM Press.

Zhu, B., & Chen, H. (2008). Communication-garden system: Visualizing a computer-mediated communication process. *Decision Support Systems*, *45*(4), 778–794. doi:10.1016/j.dss.2008.02.004