

Exploring asynchronous online discussions through hierarchical visualisation

Víctor Pascual-Cid^{1,2} and Andreas Kaltenbrunner^{1,2}

¹Universitat Pompeu Fabra, Barcelona, Spain

²Barcelona Media - Innovation Centre, Barcelona, Spain

victor.pascual@upf.edu, andreas.kaltenbrunner@barcelonamedia.org

Abstract

We introduce a highly customisable Social Visualisation system for exploring online discussions through visual representations of conversation threads. The tool provides users with an interface that allows the navigation and exploration through the often intricate structure of online discussions. Apart from being useful for readers and participants of these forums, the interactive capabilities of our system makes it appealing for social researchers interested in understanding the phenomena and intrinsic structure of online conversations. We also show a use case where we applied the tool to visualise discussions from Slashdot.org, showing its capabilities to represent new, in this case Slashdot specific, metrics. The visual representation of discussion threads has arisen as a complement for supporting investigation, as it helps to understand such large amount of information and contributes to the generation of new research ideas.

1 Introduction: Social Visualisation

Understanding patterns of human behaviour is one of the main foci of interest in sociology. In that sense, the appearance of the Internet has revealed new paradigms for analysing social behaviour, based on the study of online discussions. However, the amount of data present in these systems is usually overwhelming, reason why Information Visualisation techniques have arisen as a good approach to support social researchers.

The term “Social Visualisation” was introduced in [1] as the visualisation of social information for social purposes. According to that definition, Social Visualisation may be understood as the representation of both, the structure of the participants of a social network and their interactions within the community.

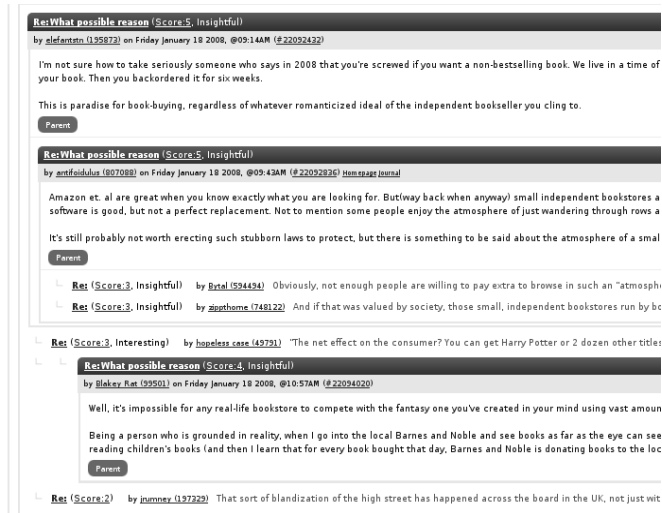
Such interactions may be found in asynchronous (blogs, wikis, newsgroups, mailing lists) or synchronous form (chats). While collecting information from chats is not an

easy task because those systems are usually proprietary and restricted, asynchronous online discussions leave a public trail that contains the contributions of the participants.

The analysis of such conversations may enable, for instance, the discovery of hot topics and appealing subjects to boost up the community. Furthermore, researchers may benefit from that information to study the behavioural patterns behind such conversations and to interpret the social aspects of online debates.

Asynchronous online discussions may be categorised depending on its structure into single-threaded and multi-threaded conversations. Single-threaded discussions are usually embedded in digital newspapers or in blogs, and provide just one thread of comments. This characteristic makes the conversation difficult to be followed by a human reader because there is no physical structure and replies to certain comments are not easy to be found. For the same reasons, automatic systems may have to run complex content-based techniques to discover the structure of the conversations. On the other hand, multi-threaded conversations usually provide indentation or adequate titles that denotes when a comment arises as a reply of another text. Two examples can be found in Figure 1, where the left subfigure shows a screenshot of Slashdot.org and the right one an example of a SourceForge discussion panel. In this examples the reader may observe the existence of a clear structure, that makes it easier to follow the different threads that were originated from the primary post. However, large debates originate countless pages with comments cumbersome to be navigated, making difficult the understanding of a whole conversation.

Due to the large amount of information existing in online discussions, Information Visualisation techniques have been widely used to develop visual interfaces that ease the exploration and understanding of this kind of data. An example of Social Visualisation was proposed in [1], where authors introduced Chat Circles, an interface for depicting synchronous conversations that enables the discovery of the role of chat participants. The authors also presented Loom which is a system for creating visualisations of the partic-



Thread View

Thread

Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

RE: Exporting images

Figure 1. Web interfaces of conversations in Slashdot (left) and SourceForge (right).

ipants and interactions in a threaded Usenet group, where posts and comments are represented as connected dots in the space. Moreover, they also classify posts according to its “mood”, generating coloured visual patterns.

Conversation Map [7] is another system designed to enable scientific and non-scientific users to visually analyse online conversations [8]. It combines the visualisation of social and semantic networks using node-link diagrams. The tool has three different frames dedicated to the visualisation of the underlying social network of a thread, its semantic relations based on discussion themes and a display area where individual threads are shown.

While previously presented tools give an overview of the discussion activity within a community, another trend has been the representation of author activity. In that sense, PeopleGarden [13] represents an original approach for depicting the behaviour of online communities. This system was based on the representation of every author in the community as a flower whose petals are ordered and coloured according to post attributes such as number of replies or posting time. Although such an organic metaphor provides an interesting overview of a community, it may fail when tackling very large communities with thousands of users.

Focusing on author activity rather than on discussion structure, [11] introduces two visualisations to support social awareness in online spaces called Newsgroup Crowds and Authorlines. These visualisations emphasise the social aspects behind the contributions of users to the community rather than the patterns that can be extracted from the discussion structures itself. Specifically, Newsgroup Crowds is a scatter plot that enables the understanding of the ac-

tivity of participants in a given newsgroup during a period of time. The plot visualises an author’s activity by depicting her number of active days and her average of posts per thread. Likewise, Authorlines represent a visualisation of the activity of a single author, showing the number of contributions to the community during one year through vertical bars made up of sized circles that represent the controversy (number of replies) generated per every author’s comments.

Finally, the impact of using visualisations to support social researchers was analysed in [12], where the authors were assisted by three visualisations (Authorlines, graph representation of ego networks and histograms of the neighbour degree distribution) to generate hypotheses on how to identify signatures of social roles.

These tools provide Social Visualisation by means of depicting author behaviour as well as the whole cloud of contributions in online discussions. However, one might miss important aspects of human behaviour when only analysing the general picture of online discussions. But surprisingly, little work has been presented regarding the representation of single conversations, perhaps with the exception of [10], where a vertical tree not suitable for representing very large debates was used. To overcome this limitations, we present here an interactive tool which allows to investigate the smallest details of conversation threads while at the same time improving the standard representation of forum threads as a whole.

Hence, the goals of our tool are both, enabling researchers to analyse a single conversation, and offering a visual interface to support discussion participants and readers to explore its contents.

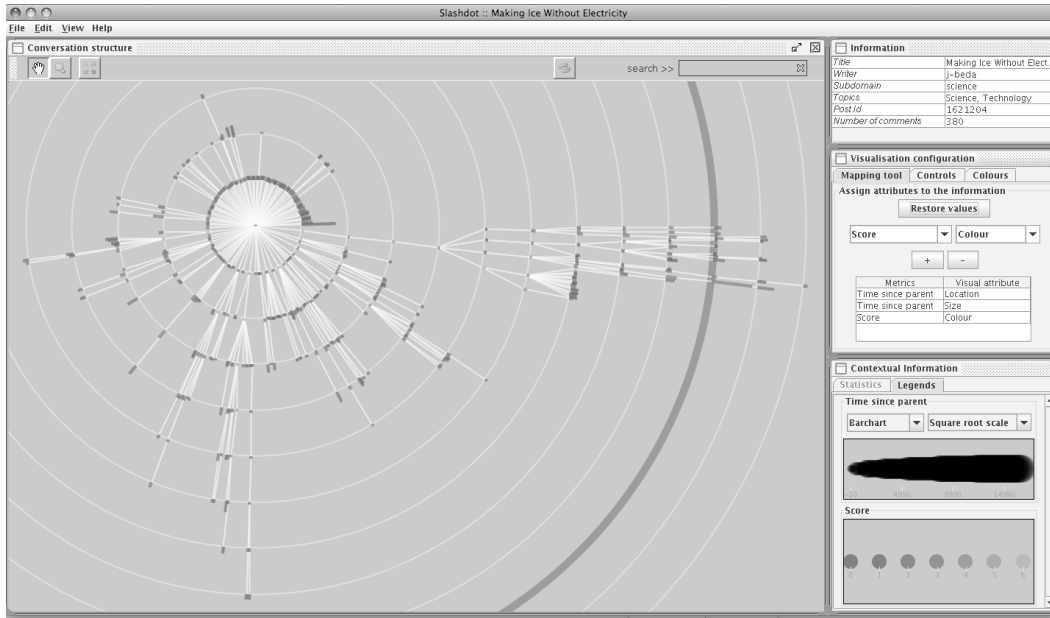


Figure 2. Discussion map of a Slashdot conversation. One dimensional bars represent posting time while shades of grey (colours in the original application) represent the score of the post.

The next section is dedicated to the presentation of our visual system for exploring online discussions, introducing its visual capabilities and browsing interactions to ease the understanding of complex discussions. Following, we will introduce a case study to show how our system can be applied to a very well known discussion site, to end up with the conclusions and future steps of our work.

2 Visualising online discussions

In [12], Welser et al. showed how visualisations assisted researchers in the analysis of social aspects of online discussions. However, as we have already seen, most of the existing visual systems address the problem of visualising all the posts and comments within a community, discouraging the analysis and navigation through single multi-threaded conversations. Such discussions may have a large amount of user contributions and represent a good source for analysing and understanding social behaviour by itself.

Typical web interfaces of newsgroups and forums use pagination and nesting as a way for enabling the navigation through multi-threaded conversations (Figure 1). Nevertheless, more complex approaches are also used where non-high rated comments are collapsed by default, leaving more screen space to valuable contributions. However, pagination and indentation are still inefficient when representing highly discussed posts with several hundreds, or even thousands of comments.

Our system uses a crawler to collect conversations from a target site, and transforms each conversation in a GraphML file used as input to the visualisation system. Such a system is an adaptation of the work presented in [6], with some design considerations based on the needs of social researchers.

Likewise, the system takes advantage of the hierarchical properties of online conversation structure, where posts may be considered as nodes and edges correspond to the replying relation between them. Hence, we use the main post as the root of a tree and its direct replies are considered part of the first level. Following replies are located in deeper levels, and so forth. The provided “discussion map” follows the Information Seeking Mantra [9], providing an overview of the discussion tree based on the Radial Tree, proposed in [2], and a set of interactions that provide details on demand. Thus, the main discussion post is placed in the centre of the tree, while its comments surround it in a concentric and nested manner, allowing the discovery of hot topics, or user-to-user debates, which can be easily identified as outliers in the hierarchy. Figure 2 represents a screenshot of the user interface of our system. While the Radial Tree depicting a conversation thread is shown within the main frame, the right part contains controls to interact with the visual attributes and legends.

The tool has been provided with the typical set of interactions like panning and zooming. Furthermore, detailed information such as author of the text, time stamp, and body of every comment are provided when hovering a node.

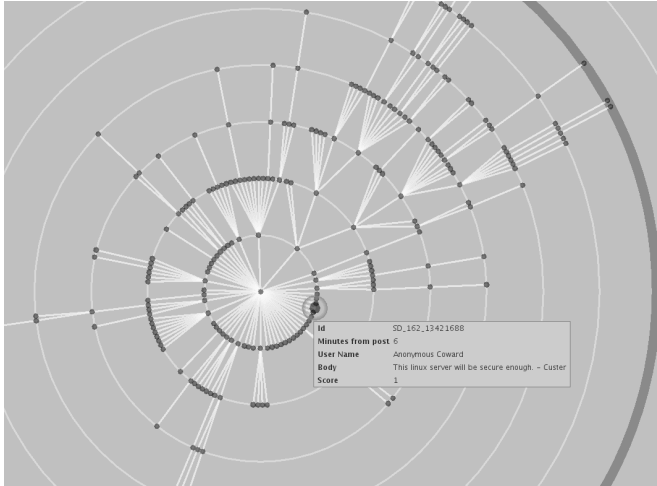


Figure 3. A tooltip information box enables users to see all the details of a comment.

Figure 3 is a snapshot of a thread with more than 200 comments, which can be read through the tooltips. The system also provides a link to every comment in the forum's original web interface.

Although the radial tree helps in building a mental map on how a discussion has evolved, it may be also of interest for sociologists or web researchers to focus their attention on a specific debate (which corresponds to a sub-tree in our hierarchy). This can be done by dragging a node to the centre of the tree. That interaction triggers an animation [14] that pleasantly relocates the nodes that are part of the sub-tree rooted at the dragged node, and hides the rest of the tree, helping the user to preserve her mental map and focusing on a sub-thread of the discussion. Furthermore, the system has a built-in search engine that enables the highlighting of posts relevant to a query defined by the user. Contrary to current search engines that show an ordered list of relevant comments, our system locates and highlights them in the discussion map, providing feedback not just on how many posts are relevant to the query, but also allowing the discovery of how many comments they have provoked.

The highlighting system also offers several modes aiming at focusing users interests and assisting in the discovery of information. For instance, it provides the highlighting of the path from a desired node to the root to help identifying a specific thread of the discussion. Another feature of this system is to illuminate the comments written by the same author, allowing to discover easily whether an author has been active in a conversation or not, or to identify all the contributions of a specific participant.

Our tool supports in-depth analysis by allowing the dynamic mapping of comment features into a set of visual attributes. This is, the user interface provides a panel where

users may select any available metric, and apply it to visual attributes of nodes in the tree. Such metrics may be defined in a configuration file, and may refer to attributes existing in the GraphML file. The system collects all the metrics and assigns its value according to the selected visual attribute, which can be one of the following:

Size: by default, the two dimensional area of the nodes is chosen to represent the values of the selected metric. However, large populated conversations may generate cluttered representations. To get rid of this problem, the system is able to represent items in one dimension, generating bars whose height depends on the metric.

Location: this visual attribute refers to the ordering of child nodes. The system locates every child in the same order as it has been visited by the crawler. However, the random nature of posts in a discussion may avoid the possibility to generate visual representations suitable to be compared across different posts. We then propose, for instance, the usage of a metric such as the total number of children per node to sort the radial tree. Hence, spiral representations may arise, facilitating the visual comparison of different threads.

Shape: different symbols may be applied to the desired metric. This visual attribute is specifically useful when dealing with ordinal or category information.

Colour: the system has two different palettes, one for quantitative values and another for ordinal or nominal ones. They can be dynamically defined by the user.

Finally, each level in the tree is depicted using a circle to reinforce the notion of depth. It also visualises a balanced depth measure in the form of a variant of the h-index h proposed in [2] as a measure of the degree of controversy of a post. This value represents the maximum nesting level i which has at least $h > i$ comments. To integrate this measure on the system, the depth corresponding to the h-index of the conversation is depicted with a red circle. When only a sub-thread of the discussion is visualised the measure is recalculated dynamically only for the corresponding sub-tree.

3 Use Case: Visualising Slashdot discussions

Slashdot is an example of a website which facilitates multi-threaded online discussions. It was created at the end of 1997 and has ever since metamorphosed into a website that hosts a large interactive community capable of influencing public perceptions and awareness on the topics addressed. The site's interaction consists of short-story posts that often carry fresh news and links to sources of information with more details. These posts incite many readers to

comment on them and provoke discussions that may trail for hours or even days. The comments can be nested when users directly reply to a comment. This way the discussions can typically reach a depth of 7 but occasionally even depths of 17 have been observed. The structure of this discussion trees has been analysed in detail in [2].

Although Slashdot allows users to express their opinion freely, moderation and meta-moderation mechanisms are employed to judge comments and enable readers to filter them by quality. The moderation system was analysed in [4], and gives an integer between -1 and 5 to every comment, being 5 the highest score.

This use case is based on news posts and the corresponding comments collected from Slashdot by a web-crawler in form of raw HTML-pages in September 2006. The collected posts were published between August, 2005 and August, 2006 on Slashdot. More details on the data and its retrieval process can be found in [3]. This raw data has been transformed into GraphML files.

We use this files as input of our visualisation tool, which then is able to represent general features of online discussions as well as some Slashdot specific metrics like comment score.

By now, our system extracts and calculates five different features per comment:

Score: an integer value between -1 and 5, indicating the comment's score obtained from Slashdot's moderation system. The mapping of this metric into suitable symbols improves the whole understanding of the hierarchy, generating a symbol hierarchy that allows to easily identify the location of top rated comments in the discussion, and to understand how debated they have been.

Time since post: represents the number of minutes since the initial post of the discussion has been published. It enables the creation of representations that provide a general feeling on how the discussion has evolved in time and allows to focus ones attention on a specific group of early or late posts.

Time since parent: similar to the Time since post, but the elapsed time is relative to the parent of the comment. This metric also allows the exploration of the evolution of the discussion in time. For instance, it may illustrate how long the users last in replying to a post.

Maximum Children's Depth and Total number of children: reinforce the notion of how much controversial a comment has been in terms of how many comments it has originated

Comment length: long comments may intrinsically incorporate more ideas, and hence, may be another interesting and easy to collect metric of a post.

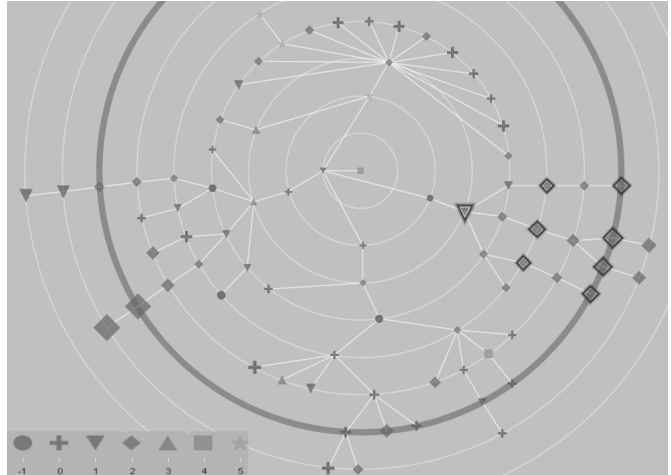


Figure 4. A sub-thread from a conversation. Score is represented with shape and colour while items size denotes comment length. All the posts from a single user in this sub-thread are also highlighted, which enables to see its participation in several different discussions.

Figure 4 shows a subthread dynamically filtered from a larger conversation. Moreover, the picture shows how shape can be applied according to a metric, in this case, the score of the comment.

Due to the visual representations based on the user highlighting and the visual representation of the h-index, researchers have been inspired to focus on the role of dialog chains in the discussions, which are far more common than expected before visualising the discussions with our tool. This chains are produced when two users start to reply to each others messages several times, often reaching a depth far greater than the h-index of the discussion.

4 Conclusions

Online discussions have arisen as one of the most common ways for getting knowledge from other users. Hence, such discussions are an interesting source for analysing human behaviour. Its hierarchical structure provides a clear organisation that enables participants and researchers to understand how a discussion has evolved. However, large conversations are difficult to be followed using the typical pagination existing in forums, newsgroups and so forth.

In this paper we have introduced a system for visualising online discussions based on their hierarchical representation. Our tool aims at enabling readers, conversation participants and researchers to navigate through the often intricate discussion structure, allowing the customisation of the vi-

sual items according to a set of comment features. The main benefit of our system is the creation of an interactive map of the conversation that may help understanding on how the discussion has evolved or how controversial it has been.

Our system aims at contributing to the discipline of Social Visualisation by complementing existing tools, offering visual approaches to analyse a discussion in detail. In that sense, we have shown a use case from Slashdot, providing some preliminary conclusions that show the benefits of the exploratory capabilities and visual mappings of our tool, as well as the discussions map, which enables final users to get an understanding on the whole conversation.

Our tool may also be easily adapted to permit the representation of a conversation in real time, being an interface for enhancing the engagement of users to the discussion.

The tool is especially useful for researchers working in the analysis of social interaction in online forums. Visual inspection of the forum conversation is very important when identifying new phenomena or finding interesting relations between structure and variables. This task is often cumbersome using standard visualisation tools our just numerical analysis. Once identified a certain interesting characteristic with our tool in specific discussions, an exhaustive analysis can be performed to proof the generality of the findings.

Our main hypothesis is that the discussion map combined with the exploratory capabilities of our system may improve the readability of the conversation as it depicts the whole conversation in the same display, facilitating the understanding of the impact of every comment and allowing the access to hot discussions and relevant posts according to user interests. The future steps of our work will include an exhaustive evaluation to prove our hypothesis, and to discover how the system may influence the behaviour of readers and participants while navigating through discussions.

Furthermore, refining the approach presented in [10] where the entire user relations network is visualised, we will include the representation of ego-networks and other user interactions in our platform. This will help to avoid the usual cluttered representations of large social networks, and still provide a general picture of user relations. The final goal of our research is to provide researchers with a full solution for analysing online communities

A demo of our tool can be found in [5].

5 Acknowledgements

This work is partially supported by the Grant TIN2006-15536-C02-01 of the Ministry of Science and Innovation of Spain and Càtedra Telefónica de Producció Multimèdia.

References

- [1] J. Donath, K. Karahalios, and F. B. Viégas. Visualizing conversation. In *HICSS '99: Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 2*, page 2023. IEEE Computer Society, 1999.
- [2] V. Gómez, A. Kaltenbrunner, and V. López. Statistical analysis of the social network and discussion threads in slashdot. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 645–654. ACM, 2008.
- [3] A. Kaltenbrunner, V. Gómez, A. Moghnieh, R. Meza, J. Blat, and V. López. Homogeneous temporal activity patterns in a large online communication space. *IADIS International Journal on WWW/INTERNET*, 6(1):61–76, 2008.
- [4] C. Lampe and P. Resnick. Slash(dot) and burn: Distributed moderation in a large online conversation space. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 543–550. ACM Press, 2004.
- [5] V. Pascual. Visualising Slashdot, 2009. <http://caw2.barcelonamedia.org/node/42>.
- [6] V. Pascual and J. C. Dursteler. Wet: a prototype of an exploratory search system for web mining to assess usability. In *IV '07: Proceedings of the 11th International Conference Information Visualization*, pages 211–215. IEEE Computer Society, 2007.
- [7] W. Sack. Conversation map: a content-based usenet newsgroup browser. In *IUI '00: Proceedings of the 5th international conference on Intelligent user interfaces*, pages 233–240. ACM, 2000.
- [8] W. Sack. Discourse diagrams: Interface design for very large-scale conversations. In *HICSS '00: Proceedings of the 33rd Hawaii International Conference on System Sciences-Volume 3*, page 3034. IEEE Computer Society, 2000.
- [9] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *VL '96: Proceedings of the 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society, 1996.
- [10] M. A. Smith and A. T. Fiore. Visualization components for persistent conversations. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 136–143. ACM, 2001.
- [11] F. Viegas and S. Smith. Newsgroup crowds and authorlines: visualizing the activity of individuals in conversational. *System Sciences*, Jan 2004.
- [12] H. T. Welser, E. Gleave, D. Fisher, and M. Smith. Visualizing the signatures of social roles in online discussion groups. *The Journal of Social Structure*, 8(2), 2007.
- [13] R. Xiong and J. Donath. Peoplegarden: creating data portraits for users. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 37–44. ACM, 1999.
- [14] K. Yee, D. Fisher, R. Dhamija, and M. Hearst. Animated exploration of dynamic graphs with radial layout. In *INFOVIS '01: Proceedings of the IEEE Symposium on Information Visualization 2001*, page 43. IEEE Computer Society, 2001.