Improving the Scalability of Interactive Visualization Systems for Exploring Threaded Conversations

A. McNutt, IEEE student member, & G. Kindlmann

Department of Computer Science, University of Chicago

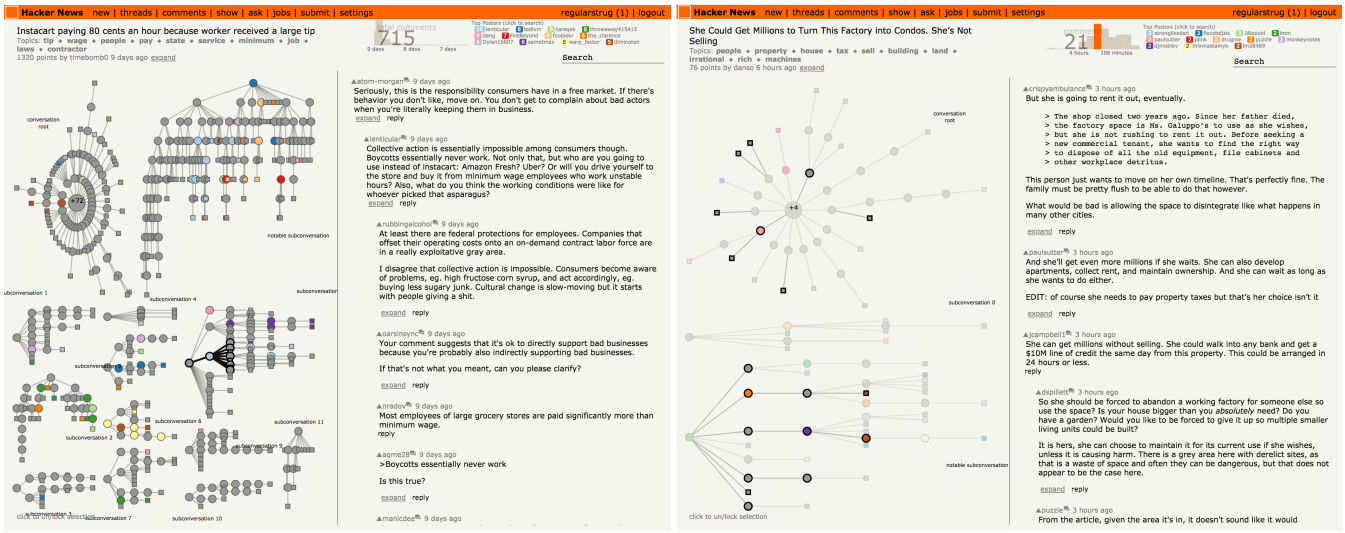

Figure 1: *Two HackerNews conversations (715 and 85 comments respectively) rendered using our ForumExplorer application. In the left image the user has moused over a particular sub-conversation; in the right they have engaged in a temporal search.*

Abstract

There has been a variety of works on developing visualization systems to augment and enhance the experience of interacting with the threaded conversations found in venues like reddit and slashdot through the use of graphical overview and nlp-generated summaries. These works have generally been designed around an ideal size of data, which can be difficult to use for large conversations, and have sometimes require non-trivial offline processing time. In tandem these factors have likely negatively affected the adoption of this class of tool. We address these problems by offering concrete design strategies that enable this type of representation to handle wider ranges of data, that we implement as a Chrome Extension, Forum Explorer, which facilitates practical real-time analysis and exploration of conversations held on yCombinator's HackerNews.

CCS Concepts

• ***Human-centered computing*** → *User interface design; Visualization; Graph drawings;*

1. Introduction

Conversation on the internet takes many shapes and forms, including question and answer forums, synchronous messaging, and asynchronous threaded conversation. Of particular interest are conversations that take place in an asynchronous threaded environments, such as reddit or slash dot, in which users are presented the opportunity to comment on the root of the conversation or on any previous comments. These forums offer a mechanism for communities to have wide collections of related conversations within a single topic. In many cases the participants in these conversations

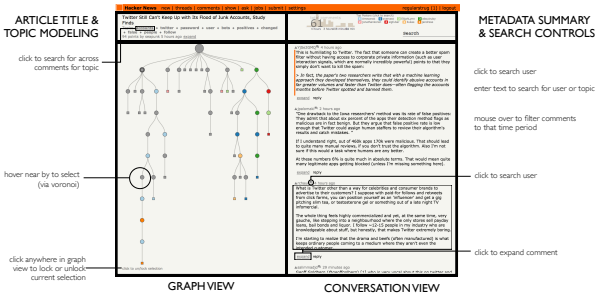

Figure 2: *Annotated view explaining the components of the UI. In addition to our Forested Tree View we support Reingold-Tilford tidy tree (both as a ring and linear tree) and gridded tree views.*

are experts on the topic and might provide valuable insights. Unfortunately the design of these digital spaces typically do not allow for users to interact with the conversational corpus as a whole, which can limit or impede understanding of the community opinions and insights about a topic.

Previous visualization works have developed a fascinating collection of UI paradigms to address this space. A common trend among these works features a overview of the conversational thread encoded as a graph-like structure, which the user then interacts with in a details-on-demand [Shn96] pattern to expose components of the discourse. While these tools are uniformly well received by their evaluation audiences, they have failed to gain widespread usage. This may be because visualization based overview systems are not well aligned with the types of tasks that people pursue on threaded forums. Optimistically we hope that visualization might provide helpful solutions, and that previous iterations have failed to gain traction because they do not possess an accessible or online implementations, are not aligned to the specific community they are trying to affect, and are designed around a single ideal size of conversation (and thus become cumbersome or difficult to use when conversations of interest fall outside of that target domain).

2. Forum Explorer

We present Forum Explorer, a Chrome Extension that repurposes the conventional layout of yCombinator's HackerNews [yCo] to facilitate better data exploration through the use of tree visualization techniques. We implement a split pane view in which simultaneously shows a graphical overview and a detail view. We provide additional detail about the UI in Figure 2. We support sub-conversational discovery by coloring vertices corresponding to the dozen top commenters, which allows for easy identification and exploration of conversations between individuals in the midst of this dialog. We provide topic model summaries of the conversation through the use of Latent Dirichlet Allocation (via lda.js [Bec16]), which we compute on a caching micro-service hosted on Heroku.

Our system expands upon previous work through a pair of novel visual encodings. Firstly, we introduce a novel Forested Tree View which splits threaded conversations at the root into a collection of smaller and more legible trees, by observing that the weights of rooted branches tends to be heavily dominate by a small collection of sub-trees. We prune the heaviest branches from the root and present them as independent trees, which organize in space by computing a SliceDice treemap consisting for a fake tree consisting of the weights of the rooted and the pruned trees. This approach allows for ample visual space to provide in-situ annotations and textual guides. We find empty space in the visualization to add these annotations (which we supply as topic model summaries) by constructing a voronoi for the total layout, and then finding the largest (and hence emptiest) cell for each subtree of the layout. Secondly, we observe that large conversations tend to be have a large number of rooted stumps that add substantial visual noise. We address this by collapsing rooted stumps into the root, while still making those nodes accessible through a details on demand interaction.

3. Related Work

Previous work on this topic has developed a number of different strategies for displaying and facilitating novel user interactions under a variety of different design goals. The first work of this type appears to be Donath et al's Loom, which introduced the idea of graphical exploration of comment graphs [DKV99]. This was followed by Sack's Conversation Map which represented conversation as a tree-like structure [Sac00]. Watternberg et al present a pair of papers which introduced split pane view, with one providing the graphical overview (which mirrored the multiply-indented form that threaded conversation are usually depicted in) and the other displaying comments [WM03, DWM04]. Pascual-Cid et al introduce a space filling radial tree layout [PCK09]. Narayan construct tldr which focuses on reddit and encodes the tree as an icicle diagram [NC10]. Hoque et al's family of tools breaks from purely metadata visualization by adding topic modeling and sentiment analysis on top of a split pane interaction with an abstracted indentation view [HC14,HC16]. Each of these tools provide valuable insight, but come at the cost of forcing their users to use an unfamiliar environment, as well as being not being actually deployed. Most closely related to our technical approach is Treeverse, a chrome extension which allows users to visualize the conversation tree associated with a particular tweet [Bul]. Our work improves over this design in that it tightens the response interaction response cycle and enables the user to extract useful information from a single view, as well as being better matched with the domain target audience.

4. Conclusions & Future Work

We have presented Forum Explorer, a tool for exploring threaded conversations on HackerNews. Our primary contribution is a novel Forested Tree layout that facilitates better scalability for systems of this type, however we believe that there is substantial room for our tool to make constributions in the future. Previous studies of threaded conversations usefully demonstrate the general usability of the constituent structural design elements. Yet it is unclear from their results whether or not these results are due to the novelty of the systems, and therein experiment bias (as noted by Isenberg et al is a frequent result of the visualization communities user studies [IIC*13]), or the true usefulness of the tool. The best evidence that this type of system has substantive utility (beyond Treeverse's modest popularity, as evidenced by 2412 active installs at the time

of this writing) is that it used by Rao et al [RS18] to study the way that twitter is used within siloed domain specific knowledge sharing discussions. Our application is well positioned, as it operates in real time, to conduct a longitudinal study that could describe might interact with this type of tool when they are able to incorporate it into their day to day workflow. The strategies described here continue the on-going dialog in the vis community on this topic, but we believe that they could also have applicability to systems outside of threaded conversations, such as visualizations of the scholarly citation graph.

References

[Bec16] BECKER K.: lda. https://github.com/primaryobjects/lda, 2016. 2

[Bul] BULTER P.: Treeverse. https://treeverse.app/. 2

[DKV99] DONATH J., KARAHALIOS K., VIEGAS F.: Visualizing conversation. *Journal of computer-mediated communication 4*, 4 (1999), JCMC442. 2

[DWM04] DAVE K., WATTENBERG M., MULLER M.: Flash forums and forumreader: navigating a new kind of large-scale online discussion. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (2004), ACM, pp. 232–241. 2

[HC14] HOQUE E., CARENINI G.: Convis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 221–230. 2

[HC16] HOQUE E., CARENINI G.: Interactive topic modeling for exploring asynchronous online conversations: Design and evaluation of convisit. *ACM Transactions on Interactive Intelligent Systems (TiiS) 6*, 1 (2016), 7. 2

[IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MÖLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2818–2827. 2

[NC10] NARAYAN S., CHESHIRE C.: Not too long to read: The tldr interface for exploring and navigating large-scale discussion spaces. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (2010), IEEE, pp. 1–10. 2

[PCK09] PASCUAL-CID V., KALTENBRUNNER A.: Exploring asynchronous online discussions through hierarchical visualisation. In *Information Visualisation, 2009 13th International Conference* (2009), IEEE, pp. 191–196. 2

[RS18] RAO S., STEIN R.: To journey in twitter canoes: Methods to understand the mechanisms and meaning in twitter conversations. In *MW18: Museums and the Web 2018* (2018). 2

[Sac00] SACK W.: Conversation map: An interface for very large-scale conversations. *Journal of Management Information Systems 17*, 3 (2000), 73–92. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996), IEEE, pp. 336–343. 2

[WM03] WATTENBERG M., MILLEN D.: Conversation thumbnails for large-scale discussions. In *CHI'03 extended abstracts on Human factors in computing systems* (2003), ACM, pp. 742–743. 2

[yCo] YCOMBINATOR: HackerNews. https://news.ycombinator.com/. 2