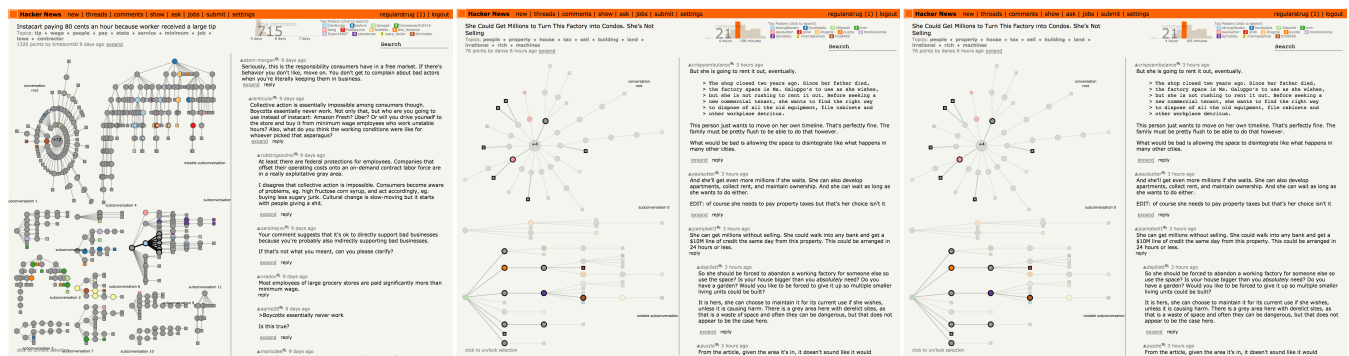


# Improving the Scalability of Interactive Visualization Systems for Exploring Threaded Conversations

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

Department of Computer Science, University of Chicago



**Figure 1:** Three HackerNews conversations (left to right 715, 85, and X) 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

The graphical presentation of large threaded conversations, such as those found on reddit and slashdot, presents individual comments well, but can cloak larger trends or patterns within the conversational corpus. Previous works have addressed this problem through visualizations and nlp-generated summaries. These works have generally been designed around an ideal size of data, which can be difficult to use for particularly large or nested conversations, and have sometimes require non-trivial offline processing time, which makes them impractical for day to day usage. We refine these approaches 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 exploration of conversations held on yCombinator's HackerNews.

## CCS Concepts

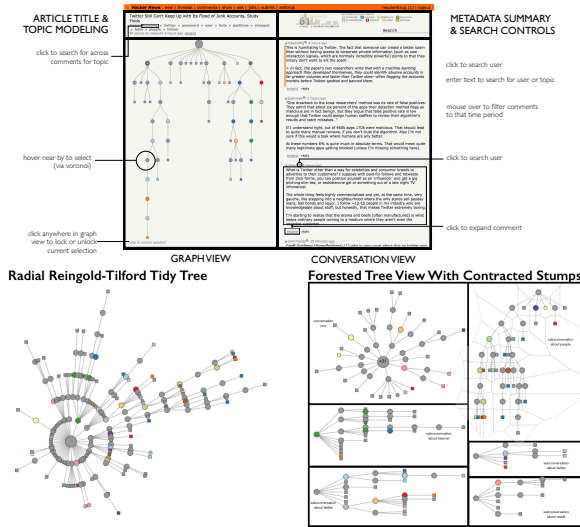
• **Human-centered computing** → User interface design; Visualization; Graph drawings;

## 1. Introduction

Conversation on the internet takes many shapes and forms. Of particular interest are asynchronous threaded conversations, such as those found 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 ranging collections of related conversations within a single topic. In many cases the participants in these conversations 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. STATEMENT ABOUT ENJOYMENT AND TASK.

Previous works have developed a fascinating collection of UI paradigms to address this space. A common trend among them features an 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 (such as groups or chains of comments). 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 fol-



**Figure 2:** Annotated view explaining the components of the UI (top) and a visual explanation of Forested Tree Layout (bottom).

lowed by Sack’s Conversation Map which represented conversation as a tree-like structure [Sac00]. Wattenberg et al introduce a split pane view, one pane providing a graphical overview (which mirrors the multiply-indented form that threaded conversation are usually depicted in) and the other displaying currently selected comments [WM03, DWM04]. Pascual-Cid et al later introduced a space filling radial tree layout [PCK09]. Narayan et al construct tldr which focuses on reddit and encodes the conversation as an icicle diagram [NC10]. Narayan et al construct encode reddit conversations through an icicle diagram [NC10]. Hoque et al break from purely metadata visualization by adding topic modeling and sentiment analysis [HC14, HC16]. Butler puts these ideas in practice through a chrome extension that visualizes the conversation trees on Twitter [But]. 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, it is unclear from the previous works. While this problem remains prescient we assert that that previous iterations may have failed to gain traction because they do not possess an accessible or online implementations, force their users out of their usual environment, 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 address these problems through Forum Explorer, a Chrome Extension that repurposes the conventional layout of yCombinator’s social news website, HackerNews [yCoa], to facilitate better data exploration through the use of tree visualization techniques. We focus on HackerNews because it has an active community (with more than 8.5 million comments) that tend to have a highly nested comment structure, and is a reputable source of domain specific opinions [BJMH15]. Our implementation captures many of the fea-

tures from previous state of the art, for which we provided an annotated guide in Figure 2 (top). Our central interaction involves the user mousing over the comment tree which is mediated through a voronoi of the graph vertices to determine the closest relevant point. 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-summaries of the comments in 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 two central improvements driven from observations about our particular domain of focus. **Firstly**, we introduce a novel Forested Tree View which splits threaded conversations at the root into a collection of smaller and more legible trees, see Figure 2 (bottom). We prune the heaviest branches from the root and present them as independent trees, which we arrange in space by computing a treemap layout. This technique allots each subtree an appropriate amount of screen area for the number of nodes that it contains and provides a helpful responsiveness for the layout. We render the root as a radial tree in order to give it visual significance, and the rest of the subtrees as linear trees whose direction (left-right vs up-down) are aligned with the longer of it’s containers dimensions (all of which are reingold-tilford tidy 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-summaries for that subtree) by constructing a voronoi for the complete layout, and then finding the largest (and hence emptiest) cell for each subtree. **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 these stumps into the root and adding a textual annotation to root indicating their presence, which allows the user to still interact with them via our details on demand pattern. Our implementation manages to scale well even to the largest HackerNews thread [yCob]. **Together** these features provide the user a rich set of overview tools that are uncluttered by matters of scale

## 3. Conclusions & Future Work

Our primary contribution in this work is a novel Forested Tree layout that facilitates better scalability in exploring threaded conversations, however we believe that there is substantial room for our tool to make contributions in the future. From previous studies it remains unclear whether or not this type of application has long-term utility (as opposed to laboratory based novelty [IIC\*13]). The best evidence that this type of system has substantive usefulness (beyond Treeverse’s modest popularity, as evidenced by its 2412 active users) is that it used by Rao et al to study the way that domain specific knowledge expressed on twitter is deeply siloed [RS18]. The practical usability of our application makes it well positioned to conduct a longitudinal study that could determine whether or not this tool would improve the effectiveness of task completion for realistic user settings. Finally we believe these design strategies 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
- [BJMH15] BARIK T., JOHNSON B., MURPHY-HILL E.: I heart hacker news: expanding qualitative research findings by analyzing social news websites. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (2015), ACM, pp. 882–885. 2
- [But] BUTLER P.: Treeverse. <https://treeverse.app/>. 2
- [DKV99] DONATH J., KARAHALIOS K., VIEGAS F.: Visualizing conversation. *Journal of computer-mediated communication* 4, 4 (1999), JCMC442. 1
- [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. 1
- [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
- [yCoa] YCOMBINATOR: HackerNews. <https://news.ycombinator.com/>. 2
- [yCob] YCOMBINATOR: HackerNews: Google Fires Employee Behind Controversial Diversity Memo. <https://news.ycombinator.com/item?id=14952787>. 2