

Improving the Scalability of Interactive Visualization Systems for Exploring Threaded Conversations

A. McNutt and G. Kindlmann

Department of Computer Science, University of Chicago

Abstract

Large threaded conversations, such as those found on YCombinator's HackerNews, are typically presented in a way that shows individual comments clearly, but can obscure larger trends or patterns within the conversational corpus. Previous research has addressed this problem through graphical-overviews and NLP-generated summaries. These efforts have generally assumed a particular (and modest) data size, which limits their utility for large or deeply-nested conversations, and often require non-trivial offline processing time, which makes them impractical for day-to-day usage. We describe here Forum Explorer, a Chrome extension that combines and expands upon prior art through a collection of techniques that enable this type of representation to handle wider ranges of data in real time. Materials for this project are available at <https://osf.io/nrhqw/>.

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 on reddit or HackerNews, in which users comment on the root of a thread or any previous comment, thus forming a tree. Unfortunately the design of these digital spaces typically do not allow users to interact with the conversational corpus as a whole, which can limit or impede understanding of community opinions and insights about a topic. They typically offer an entry point to the thread through the first or top comment, which might not be representative of the full discussion. This suggests the utility of an overview not based on a linear ordering. Participants in these conversations might provide domain expertise or other valuable insights, which can get lost in the crowd. In fact, expertise-seeking tasks are sometimes the primary motivators for forum usage [BJMH15]. More generally, user tasks in these forums manifest themselves as “discover” and “browse” tasks [BM13, HC14].

Prior research developed a collection of UI paradigms to augment and enhance threaded forums. They frequently feature an overview of conversational threads encoded as graph-like structures, which the user interacts with in details-on-demand patterns [Shn96] and through scented-widgets [WHA07] to expose components of the discourse (such as chains of comments). Donath et al. [DKV99] introduced the idea of graphical exploration of comment graphs, which Sack [Sac00] organized into a tree-like structure. Wattenberg et al. [WM03, DWM04] introduced a split-pane view, one pane providing a graphical overview (which mirrors the

multiply-indented form that threaded conversations are usually depicted in) and the other displaying currently selected comments. Pascual-Cid et al. [PCK09] later introduced a space-filling radial tree layout. Narayan et al. [NC10] focused on reddit and encoded conversations as icicle diagrams. Butler [But] put these ideas into practice on Twitter conversations through a Chrome extension, *Treeverse*. Hoque et al. [HC14, HC16] broke from purely metadata visualization by adding topic modeling and sentiment analysis. Previous tools are often designed around a particular (and modest) size or shape of data, which can cause conversations that are outside of this pre-identified form to be difficult or cumbersome to understand. Further, prior work has not investigated the broader or real-world ecological validity of the proposed approaches, which would benefit from online or in-situ implementations. Such implementations either do not exist, or require non-trivial offline processing, which our approach avoids. Rao et al.'s [RS18] study of the siloing of domain specific knowledge on Twitter uses *Treeverse*, which suggests a long term study may yield positive results. In this work we develop a small collection of novel techniques and engineering solutions which address these issues and upon which future field studies can be conducted.

2. Forum Explorer

We now describe *Forum Explorer*, a Chrome extension that enables graphical exploration of yCombinator's HackerNews (HN) [yCoa]. We focus on HN because it has an active community, features a developer-friendly API, and is seen as a reputable source of

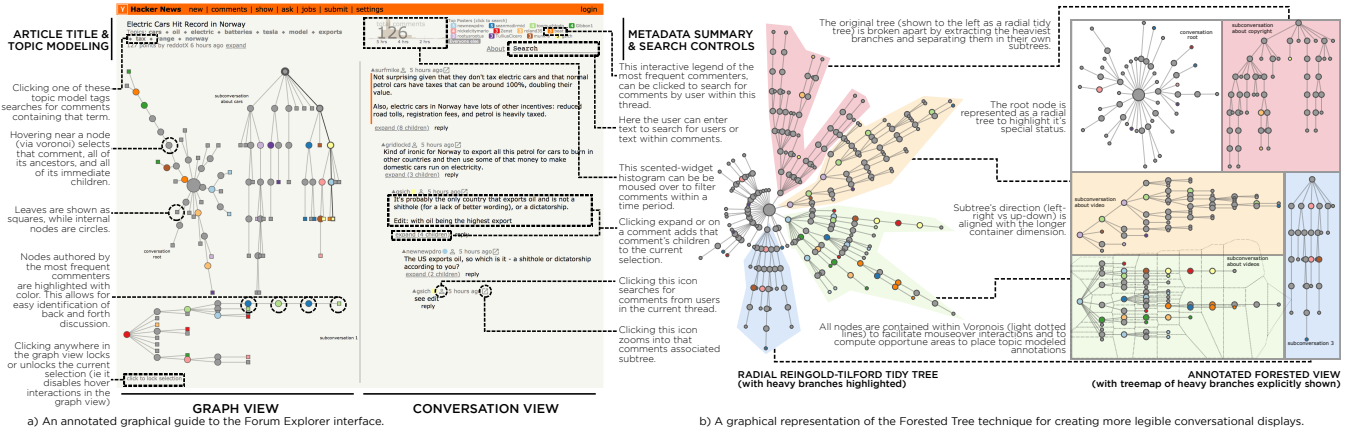


Figure 1: An annotated view of Forum Explorer (a) and a graphic description of the forested tree view (b).

domain-expert opinions [BJMH15]. Our implementation captures many of the features from prior art, including a tree-based graphical overview that denotes comments as vertices and parentage as edges, keyword and free-text-based searches, and NLP-based summaries. We offer summaries of the conversation through Latent Dirichlet Allocation topic models (via *lda.js* [Bec16]) which we present as clickable tags summaries (which are shortcuts to keyword searches). We annotate our design in Fig. 1a.

Our design is motivated by three observations about the particularities of our problem domain. **First**, we observe that the weights of rooted branches tend to be heavily dominated by a small collection of subtrees. To this end, we introduce a novel *Forested Tree View* which prunes the heaviest branches and presents them as independent trees arranged, with the root, as a squarified treemap [BHVW00]. This technique combines previous graph-based approaches [HC14, PCK09] with treemap-based techniques [FS02, EKM05] in a manner that preserves the granular detail of the graph displays while gaining the flexible layout of the treemap. We detail this process in Fig. 1b. This technique allots each subtree an appropriate amount of screen area for the number of comments that it contains. Using treemaps as a layout system endows the system with an easily recalculable responsive design strategy. The root is rendered as a radial tidy tree in order to give it visual significance, and the rest of the subtrees as linear tidy trees whose direction (left-right vs up-down) are aligned with the longer container dimension [RT81]. Available javascript treemap implementations do not support location constraints as part of layout, but the smooth animation transition helps disambiguate correspondences between subtrees in different views. We find empty space in the graphic to add annotations (which are single 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. **Second**, we observe that large threads tend to have a large number of comment stumps on the root, which adds substantial visual noise. We address this by collapsing the stumps and adding a textual annotation to the root. The user is able to interact with these hidden nodes by clicking the root and browsing the conversation view. **Third**, we observe that reading threads is pseudo-fractal, in that each subtree is at least somewhat understandable independently. To this end we

enable users to zoom into and out of the tree by providing links that transform the current layout to one rooted at the targeted node. If a user then encounters an exceptionally large tree, this mechanism allows them to access a comprehensibly-sized subtree. **Together** these features provide a set of overview and filtering tools that facilitate both “browse” and “discover” tasks [BM13].

Our application visually scales and maintains responsiveness on even the largest HN thread [yCob]. All processing occurs in the browser in real time, except for the topic modeling computed on a cloud-based micro-service. For a thread of n comments the time complexity of the tidy-tree algorithm is $O(n)$ [BJL02] while the Voronoi computation is $O(n \log n)$ [For87]. Our profiling has shown that the current bottleneck is rendering thousands of DOM elements (other costs are mitigated through caching), which can be improved through virtualization [Vau19] and raster rendering. In addition to caching computation results, we client-side cache comment data (via IndexedDB) for fast loads. This can cause the display to be out-of-date while we execute a correcting load in the background, though due to the asynchronous nature of threads, this does not have a significant effect on user experience.

3. Conclusions & Future Work

We have presented *Forum Explorer*, a tool for exploring threaded conversations on HackerNews. Our key contribution is the *Forested tree view* which facilitates exploration of larger conversations in a visually scalable manner. We intend to refine our treemap algorithm to preserve adjacency between the separated branches; this improved spatial consistency will better match users’ natural mental model of the conversation tree. The practical usability of our application makes it well-positioned to conduct a field study that could answer questions about the real-world utility of this class of visualization systems. In future work our application could be expanded to cover other forums, like reddit, by developing text mining tools (unnecessary in this work because of the advantageous HN API). Future work might expand upon Engdahl et al.’s [EKM05] work on stylus-mediated thread exploration via treemaps by considering how to translate our treemap-plus-graph approach to touched-based mobile environments.

References

- [Bec16] BECKER K.: `lda.js`. <https://github.com/primaryobjects/lda>, 2016. 2
- [BHVW00] BRULS M., HUIZING K., VAN WIJK J. J.: Squarified treemaps. In *Data Visualization 2000*. Springer, 2000, pp. 33–42. 2
- [BJL02] BUCHHEIM C., JÜNGER M., LEIPERT S.: Improving walker’s algorithm to run in linear time. In *International Symposium on Graph Drawing* (2002), Springer, pp. 344–353. 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. 1, 2
- [BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2376–2385. 1, 2
- [But] BUTLER P.: Treeverse. <https://treeverse.app/>. 1
- [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. 1
- [EKM05] ENGDAHL B., KÖKSAL M., MARSDEN G.: Using treemaps to visualize threaded discussion forums on pdas. In *CHI’05 extended abstracts on Human factors in computing systems* (2005), ACM, pp. 1355–1358. 2
- [For87] FORTUNE S.: A sweepline algorithm for voronoi diagrams. *Algorithmica* 2, 1-4 (1987), 153. 2
- [FS02] FIORE A. T., SMITH M. A.: Interactive poster: Treemap visualizations of newsgroups. In *Interactive Poster at IEEE Symposium of Information Visualization, Boston, Massachusetts* (2002). 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. 1, 2
- [HC16] HOQUE E., CARENINI G.: Interactive topic modeling for exploring asynchronous online conversations: Design and evaluation of Con-VisIT. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 6, 1 (2016), 7. 1
- [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. 1
- [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. 1, 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). 1
- [RT81] REINGOLD E. M., TILFORD J. S.: Tidier drawings of trees. *IEEE Transactions on software Engineering*, 2 (1981), 223–228. 2
- [Sac00] SACK W.: Conversation map: An interface for very large-scale conversations. *Journal of Management Information Systems* 17, 3 (2000), 73–92. 1
- [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
- [Vau19] VAUGHN B.: `react-virtualized`. <http://www.reactvirtualized.com>, 2019. 2
- [WHA07] WILLETT W., HEER J., AGRAWALA M.: Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1129–1136. 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. 1
- [yCoa] YCOMBINATOR: HackerNews. <https://news.ycombinator.com/>. 1
- [yCob] YCOMBINATOR: HackerNews: Google Fires Employee Behind Controversial Diversity Memo. <https://news.ycombinator.com/item?id=14952787>. 2