Towards an Exploratory Visual Analytics System for Multivariate Subnetworks in Social Media Analysis

Zeyang Huang*
Linköping University

Kostiantyn Kucher†
Linköping University
Linnaeus University

Andreas Kerren ‡
Linköping University
Linnaeus University

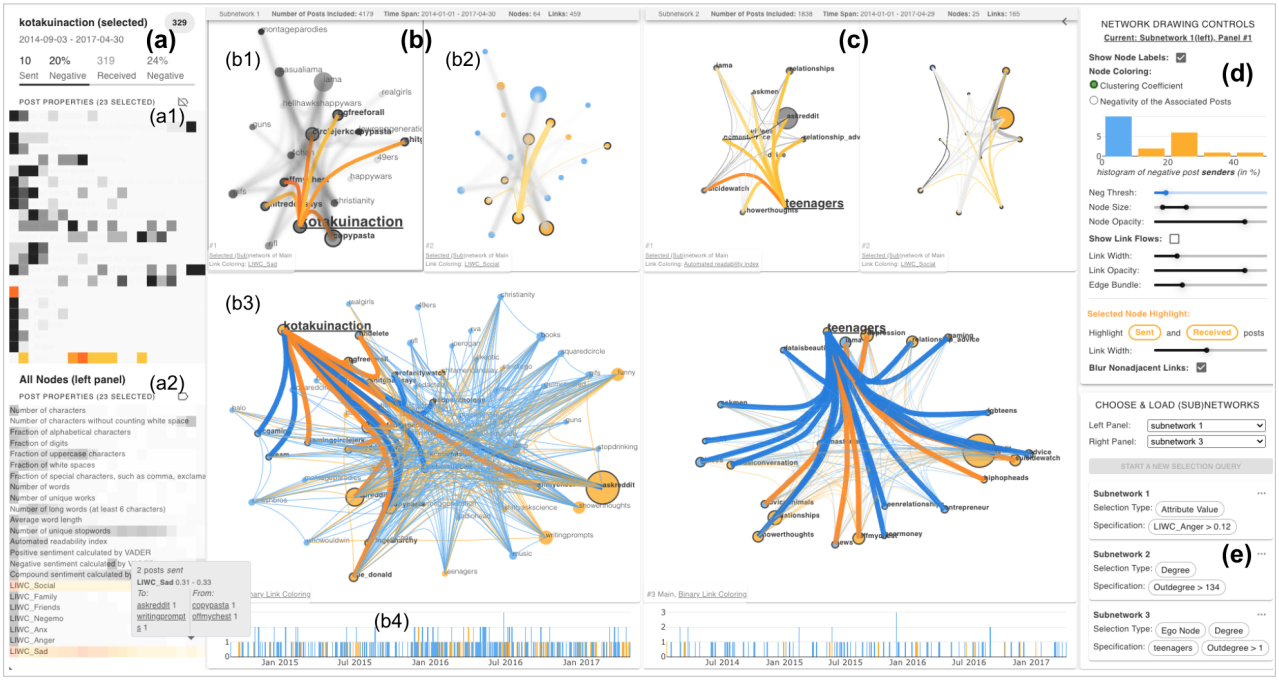

Figure 1: A screenshot of our interface with (a) the summary statistics panel and two pixel-based attribute frequency histograms for a selected node and its corresponding subnetwork; (b) the main panel of one selected subnetwork, consisting of (b3) a node-link diagram for the complete subnetwork, (b1) and (b2) two views displaying different aspects of the current subnetwork with user-defined color schemes, and (b4) a timeline view of all posts belonging to the current selection; (c) the main panel of another independently chosen subnetwork for horizontal, parallel visual comparison; (d) the network drawing control panel for each node-link diagram; and finally (e) the network selection panel reflecting current choices of displayed subnetworks and their selection criteria.

Abstract

Identifying sociolinguistic attributes of inter-community interactions is essential for understanding the polarization of social network communities. A wide range of computational text and network analysis methods may be applicable for this task, however, interpretation of the respective results and investigation of particularly interesting cases and subnetworks are difficult due to the scale and complexity of the data, e.g., for the Reddit platform. In this poster paper, we present an interactive visual analysis interface that facilitates network exploration and comparison at different topological and multivariate attribute scales. Users are able to investigate text- and network-based properties of social network community interactions, identify anomalies of conflict starters, or gain insight into multivariate anomalies behind groups of negative social media posts.

*e-mail: zeyang.huang@liu.se
†e-mail: kostiantyn.kucher@{liu, lnu}.se
‡e-mail: andreas.kerren@{liu, lnu}.se

Poster Abstract, IEEE Visualization and Visual Analytics (VIS '22), Oklahoma City, USA (Hybrid), 2022.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools; Human-centered computing—Visualization—Visualization application domains—Visual analytics

1 Introduction

Social media has been increasingly polarized over the years, yet inter-community interactions often trigger conflicts that lead to homophily and further polarization. If we examine the language used inside each user-defined community, such as Reddit user groups, using the existing text analysis methods such as LIWC [7], for instance, the linguistic properties of posts would reveal the ideological segregation and social relationships of the corresponding user groups. Prior studies have also shown that the majority of internet conflicts were triggered by a small subset of communities [4].

Exploratory analysis of online community behaviors has been widely studied in the field of Social Network Analysis (SNA) [2]. For network data with a wide spectrum of attributes and labels, visualizing multivariate information and network structure together with a good balance still remains challenging [3,6]. Given the SNAP subreddit hyperlink network data set [4], which contains posts that have hyperlinks from one subreddit community to another, along

with rich multivariate and temporal information at a large scale, the need to scale down the network and compare multiple subnetworks with different encoded information at once arises. Based on this data set, we propose an exploratory visual analytics tool for subnetwork comparison at various structural and multivariate attribute scales, to help users gain insight into polarization patterns and lexical attribute anomalies behind complex social media interactions.

2 System Design

Our visualization tool facilitates users in investigating and comparing subnetworks of a given large-scale social network data set (the original data set by Kumar et al. is based on 40 months of data [4], and thus our approach is not designed to represent the complete network). Users are able to gain insights into structural and topological differences and detect multivariate attribute patterns between and within each of the subnetworks. Specifically, users may start by defining subnetworks they want to focus on, choose two of them to be displayed on the screen, explore multivariate attributes associated with each subnetwork, investigate ego nodes of interest, and customize network drawings to make new or further comparisons.

Overview of Community Polarization. The main node-link diagram section (Fig. 1(b3)) provides an overview of all the nodes and links within a selected subnetwork. Each node represents a subreddit community, and links represent the interactions between two communities. An interaction is recorded if the content of a post in a subreddit community contains a hyperlink to another subreddit. The positions of nodes correspond to 2D dimensionality reduction (t-SNE) results for node embeddings. Similar node embedding vectors mean that users who posted in each subreddit are similar [5]. Repulsion and force-directed edge bundling are introduced to alleviate visual clutter. In this main view, orange and blue colors are used to represent sentiment polarity and show the overall polarization tendency for each community and their pairwise interactions. Since a node contains posts sent and received from a subreddit community, and a link is also an aggregation of post interactions between two nodes, we specify a user-defined threshold value and apply the orange color to a node/link to indicate the average negativity. A summary histogram of all negative post senders shown in Fig. 1(d) enables users to choose a reasonable threshold value to highlight, quickly identify individual communities or interactions that created more negative posts, and select them for further investigation. Similarly, a timeline view with negative posts highlighted in orange provides additional information for temporal polarization changes.

Displaying Multivariate Attributes. The two smaller sections above the main diagram in Fig. 1(b1, b2) show a single user-defined aspect of the chosen subnetwork. The currently displayed nodes are from a user-chosen subset of the selected subnetwork. Since there are multivariate attributes associated with each post (specifically, text features such as the number of words, automatic readability indices, LIWC anxiety values, etc.), our interface facilitates direct comparisons of two multivariate attribute assignments using the different color intensity of links. The link color is darker if the selected attribute value is higher. For instance, users can assign Fig. 1(b1) with the attribute "LIWC anger", and Fig. 1(b2) with the attribute "automated readability index". Then by comparing the color intensity assigned to each link, users can gain insight into correlations or anomalies of community interactions under the consideration of those predefined attributes. Upon the selection of one node, its neighbors are highlighted and multivariate attributes are displayed in the left panel (see Fig. 1(a1)), along with the summaries of the respective subnetwork. Each row represents a multivariate attribute, and each pixel represents a histogram bar, with the intensity of colors representing how many posts are associated with the attribute within this particular range. Users can detect anomalies by hovering the pixels at the right end, and see which communities were sending or receiving such posts that contain a high attribute value.

Subnetwork Comparisons and Customization. The main screen area is divided into left and right sections, with each side representing a subnetwork. While users might find a particular node interesting, potentially they could start a new subnetwork selection query (see Fig. 1(e)) to get the entire ego network of that node from the original data set. For example, while exploring the subnetwork in Fig. 1(b), users identify an interesting community "teenagers". They can compare the original subnetwork to Fig. 1(c), which shows the ego network of this subreddit community "teenagers" in the scope of the complete data set. Our alignment panels help users in detecting network topology patterns and help tackle the issue of scalability for a full-size large-scale network. Besides mouse-dragging to change the position of nodes, in Fig. 1(d), users can both customize the visual representation, such as if the node should be colored by the clustering coefficient or the negativity of the associated posts, and adjust specific network drawing conditions such as background blur and edge bundling force to reduce potential visual clutter.

3 Conclusions and Future Work

This poster paper proposes a new exploratory visual analytics system to facilitate comparisons between multiple subnetworks at different structural scales and from different multivariate aspects. It provides an aggregated overview for users to gain insights into the topological and multivariate textural properties of large-scale social network interaction data. Users can compare extracted subnetworks and detect similarities, attribute anomalies, and ego nodes of interest.

The system is still in the early development stage. We plan to extend our work by adding interactive self-defined subnetwork selection queries with the help of subgraph selection techniques and natural language processing tools. We also plan to incorporate more interactions for temporal aspects [1] and visualize network evolution features [8]. Furthermore, we plan to improve the overall user interface experience by optimizing the existing panels' rendering process for smoother transitions, incorporating original text data as detail on-demand, and performing user studies to evaluate our system with domain experts, including scenarios with other data sets.

Acknowledgments

This work was supported through the ELLIIT environment for strategic research in Sweden.

References

[1] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum*, 36(1):133–159, Jan. 2017. doi: 10.1111/cgf.12791

[2] M. Freire, C. Plaisant, B. Shneiderman, and J. Golbeck. ManyNets: An interface for multiple network analysis and visualization. In *Proc. CHI*, pp. 213–222. ACM, 2010. doi: 10.1145/1753326.1753358

[3] A. Kerren, H. Purchase, and M. O. Ward, eds. *Multivariate Network Visualization*. Springer, 2014.

[4] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky. Community interaction and conflict on the web. In *Proc. WWW*, pp. 933–943. IW3C2, 2018. doi: 10.1145/3178876.3186141

[5] S. Kumar, X. Zhang, and J. Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proc. KDD*, pp. 1269–1278. ACM, 2019. doi: 10.1145/3292500.3330895

[6] C. Nobre, M. Meyer, M. Streit, and A. Lex. The state of the art in visualizing multivariate networks. *Computer Graphics Forum*, 38(3):807–832, June 2019. doi: 10.1111/cgf.13728

[7] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, Mar. 2010. doi: 10.1177/0261927X09351676

[8] L. Xie, J. O'Donnell, B. Bach, and J.-D. Fekete. Interactive time-series of measures for exploring dynamic networks. In *Proc. AVI*, pp. 1–9. ACM, 2020. doi: 10.1145/3399715.3399922