

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

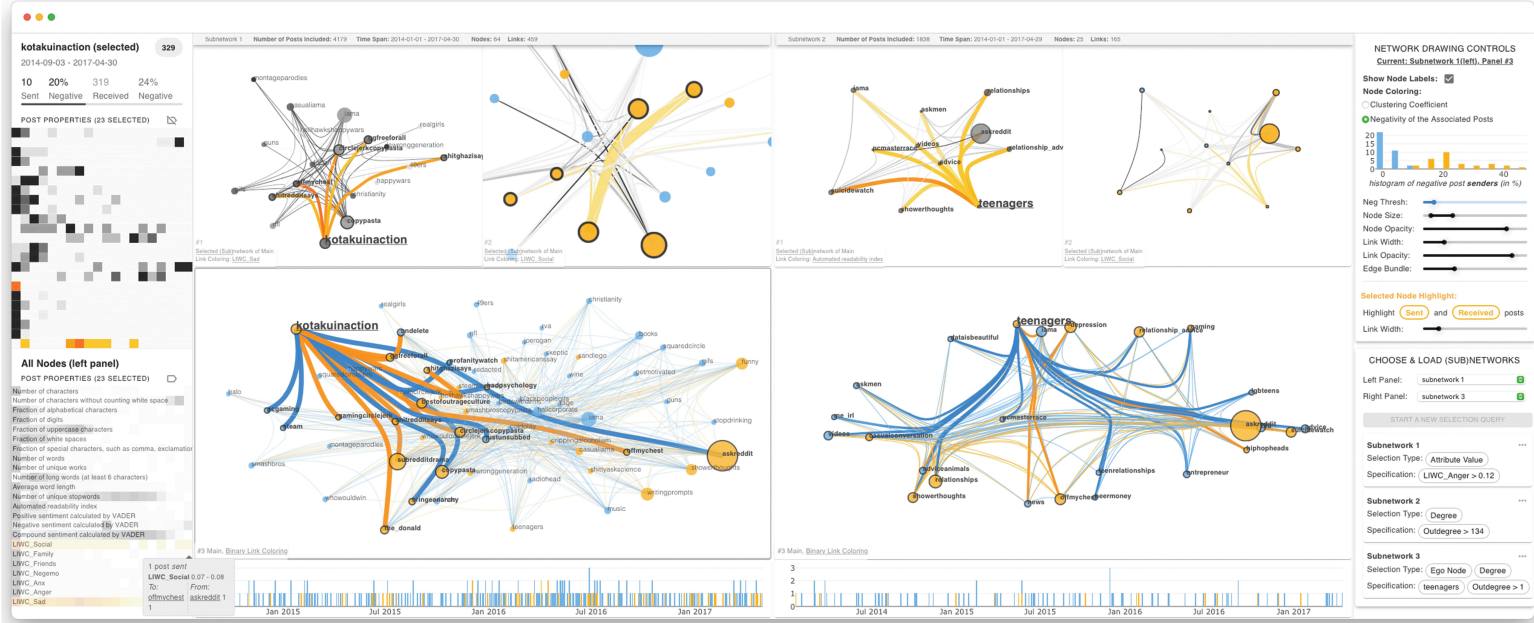
Zeyang Huang*
Linköping University

Kostiantyn Kucher
Linköping University
Linnæus University

Andreas Kerren
Linköping University
Linnæus University

*Contact:
zeyang.huang@liu.se

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.



A summary statistics panel
with pixel-based attribute frequency histograms
(for a selected node)
show each post's multivariate attributes e.g., word count, readability index, LIWC categories, ...
gain insight into correlations or anomalies of community interactions
(for the selected node's corresponding subnetwork)
assign to b1/b2/b3/...
min_attribute_value max
fewer posts in this range
more posts

B main panel 1: subnetwork 1
show a single user-defined aspect of the chosen subnetwork e.g., a user-chosen subset of the selected subnetwork in panel B
b1 aspect 1
b2 aspect 2
use different color intensities to compare attribute value e.g., assign links in b1 with the attribute LIWC anger, and b2 with the attribute automated readability index
overview of all the nodes and links in a subnetwork
The displayed central view above uses a binary orange and blue color encoding, aiming to show overall polarization tendencies for each community and their incoming and outgoing interactions. Since the polarity value is aggregated by a large number of posts, users can refer to the polarity distribution in panel D, and choose a desired negative highlight (orange) threshold value.
recommended workflow E → B & C → A → E
b3
b4 timeline: post distribution from 2014 to 2017, colored by binary sentiments

C main panel 2: subnetwork 2
the main panel of another independently chosen subnetwork for horizontal, parallel visual comparison
node: a subreddit community, from the SNAP subreddit hyperlink network dataset
node size: number of posts associated with this subreddit
node position: t-SNE results for node embeddings, close if users that post in the subreddits are similar
node color: user-defined color schemes
option 1: aggregated sentiment polarity (blue or orange)
option 2: clustering coefficient (color intensity)
link: aggregation of posts, showing pairwise interactions between two communities*
link width: post inflow, outflow, or both in this interaction
link position: repulsion force and adjustable edge bundling force
link color: user-defined color schemes
option 1: the selected multivariate attribute (color intensity)
option 2: binary sentiment classification (blue or orange)
neutral or positive negative small to large

D network drawing control panel
for each selected node-link diagram
control visual representation e.g., coloring schemes
adjust the drawing of each node-link diagram e.g., edge bundling force

E network selection panel
display subnetwork selection criteria
give overview of network properties

We propose a visual analytics tool that provides:

- ✓ Overview of community polarization (panel B & C)
- ✓ Displaying multivariate attributes (panel A)
- ✓ Subnetwork comparisons and customization (panel D & E)

Our approach is not designed to represent the complete data set. Instead, we try to tackle the issue of scalability for a full-size large-scale network, by zooming into areas of interest from multivariate, topological, and temporal aspects. We aim to provide users with an exploratory visual journey in a large-scale social network data set like this, and facilitate comparisons and selections of multiple subnetworks from different perspectives.

Ongoing and Future Work

- ✎ Add interactive self-defined subnetwork selection queries with the help of subgraph selection techniques and NLP tools
- ✎ Incorporate more interactions for temporal aspects and visualize network evolution features
- ✎ Optimize performance
- ✎ Add original text data for details on demand
- ✎ Perform user studies to evaluate our system with domain experts, including scenarios with other data sets.