CSE 578: Data Visualisation

R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media

Parv sethi
CIDSE
Arizona State University
Tempe, Arizona

Anay Jain
CIDSE
Arizona State University
Tempe, Arizona

Aditya Rohilla
CIDSE
Arizona State University
Tempe, Arizona

Harika Kondru
CIDSE
Arizona State University
Tempe, Arizona

Sheen Dullu
CIDSE
Arizona State University
Tempe, Arizona

ABSTRACT

The R-Map tool we implemented is an extension to the existing visual analytical system, R-Map[1], which supports the reposting structure and allows to interactively explore and analyze the information in the social media. A single original media post from the right person or say a key player can create abundant repostings which can involve thousands of people with different opinions. For better understanding, we used map metaphor to describe nodes and relations are represented by key players, rivers, routes, and bridges in the virtual geographical space. Additional features on the tool are provided to investigate user sentiments, patterns and comparing key players' volume of reposts. We further evaluated the effectiveness and usability of the R-Map system with two datasets and their case studies.

I. INTRODUCTION

People use social media to share their thoughts or information to other users, making social media a vital platform for users. These platforms allow replying to messages or reposting the original message to help spread the message on the social media platform. Through this reposting process, the key players such as politicians or entertainers may cause a large amount of reposting due to their millions of followers making them influential leaders in social media.

R-Map focuses on exploring the diverse reposting cascades with different semantics and multiple levels. The system mainly focuses on multi-level structure and semantics of reposting behavior. To effectively visualize multi-levels reposting structure, R-Map uses a map metaphor to help communicate the reposting pattern of the initiated tweet in a spatial context. Terminologies such as lakes is referred to as key players and users that repost the key player's message or post are named counties. Similarly, for relations, the reposting relationship between key players is represented by rivers. Our project is based on understanding and implementing the R Map system, also adding some new extensions to improve the functionalities of the system.

II. VISUALIZATION DESIGN

2.1 Data Attributes

The data in the original R-Map system is taken from the social media platform named Sina Weibo which is widely used in China. However, our data is from Twitter social media. Both the datasets are similar as both follow the tweet structure and reposting structure. Analyzing the reposting process of an original tweet/weibo is the primary focus of the R-Map system. The following

terminologies are used to define the reposting data:

- The root user is the originator of the original tweet/weibo.
- Reposting: the direct and indirect repostings by the followers of the root user.
- Non-follower repostings are tweets/weibos that are reposted by users who do not follow the root user.
- Key players are the social media handlers who produce large-scale repostings to the original tweet/weibo.

A large amount of user repostings forms a tree-like structure and the original post is referred to as a root node.

2.2 System Interface

The system includes a Map View, a Table View, a Word Cloud View and a Timeline View that supports the visual exploration. The overview of the reposting structure and process is shown in the Map View. It consists of visual metaphors such as countries, regions, lakes, etc. Table View provides a list of all weibos. Users can sort them according to reposting time, number of repostings, etc. Clicking and selecting any weibo gives an information panel for the reposting user. The Word Cloud View gives a keyword distribution of the weibo. The size of any word in the word cloud denotes its frequency. The Timeline View presents the whole diffusion process in a temporal overview. The timeline shows the delay time of reposts from the beginning, i.e., created time of the original weibo/tweet.

Due to differences in datasets and the characteristics of data between Twitter and Weibo, we implemented a few features from the original system and added extensions on it. In our system, the Table View provides a list of 20 tweets from the root users along with information such as root user name, Tweet ids, and time of the creation of tweets. We allow users to sort the table according to username and tweets. By clicking on a tweet, the Word Cloud and Timeline View get created and are presented with data. Word Cloud consists of all the keywords in the reposting network of that tweet. The Timeline View depicts the number of repostings for a tweet that is, it lets the user know the rate at which the retweets on the tweet disperses. The timeline view offers a window of 24 hours from the original post.

III. EXTENSION

While implementing the original R-Map system, we added some new features which enhance the visual analytical system and adds up to the required objective of the system.

First, analyzing the sentiment analysis of all the retweets on the original tweets of the root users. This extension helps determine what percentage of the followers or non-followers of the root users agree or disagree with the tweet. We determined the sentiment analysis of all the retweets and visualized them based on positive, negative, or neutral sentiments. We concatenated the Bubble Chart and Pie Chart to visualize this behavior. The size of the Bubble Chart corresponds to the count of retweets for a tweet of key player has, and the distribution of colors in the Pie Chart representing sentiments.

Second Funnel Chart, this feature shows size of reposts of entire root users. This feature helps us to identify the top root users whose posts helps accelerate the diffusion process. This feature is essential in the process of dispersing information to a broader audience.

For the third extension, we added a Word Cloud depicting the sentiment analysis for every key player for a root user. This Word Cloud displays the positive, negative or neutral words for every tweet.

IV. CASE STUDIES

To run the R-Map interface we created two datasets. The first dataset we used is of politicians and the other we used is of non-politician or entertainers. Due to lack of proper resources we created fake data for both datasets.

When the user selects a tweet from the Table View (Figure 1A), the packed Bubble and Pie Chart (Figure 1B), Timeline Visualization (Figure 1C) and Word Cloud (Figure 1D) is shown in the system interface.

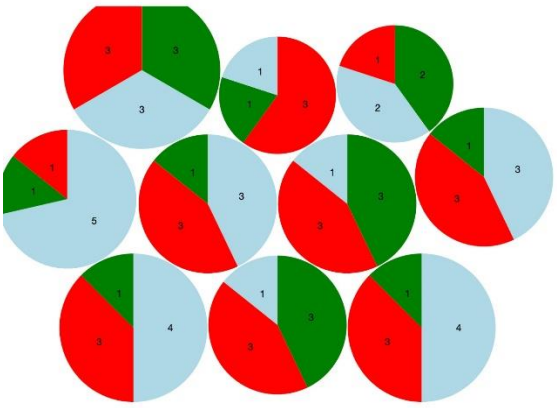

Figure 1A. Packed Bubble and Pie Chart

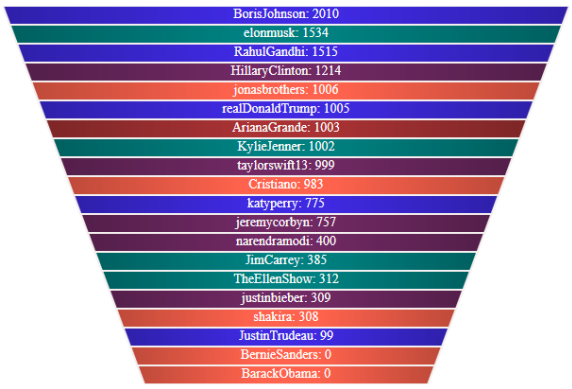

Figure 1B. Funnel Chart.

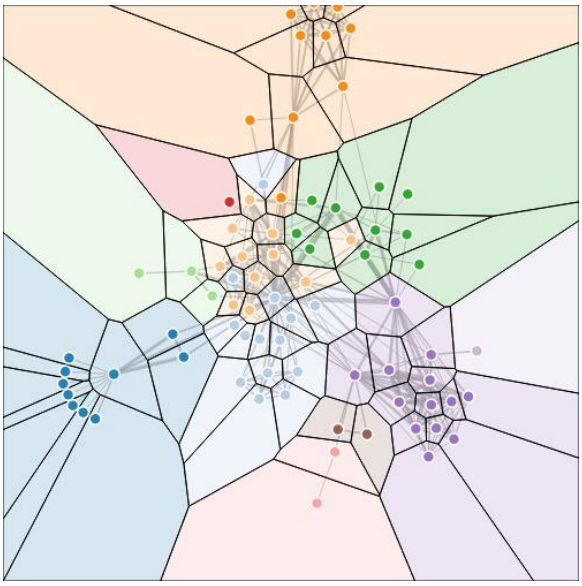

Figure 1C. R-Map

When the user clicks on the tweet from the Table View the system shows the below interface with R-Map Figure 1C. Ech color depicts a

Figure 1D. R-Map Interface

Each color depicts a different key player in R-Map Figure 1C. The color gradient denotes the level of reposting.

V. DISCUSSION AND FUTURE WORK

In addition to the social media data, R-maps can also be used to visualize hierarchical structures like company structures. The R-map can be further enhanced using accurate data from twitter. The shape of the visualization depends on the extracted key players. This process may take up more space

and look less compatible with the window if the key player's number is high. So, a constant shape and size of an island would be better in the visualization. The data is not dynamic, i.e., that visualization depends on the extracted key players. So, if the data of extracted key players changes. It is not possible to show in this visualization. Too many visualizations on a single-window give appropriate information to the user, but it forms a visual clutter so an overview and zoom on-demand filter should be added to the visualization. The sentiment analysis on tweets can be improved using state-of-the-art-natural language processing as mentioned in the base paper. The further enhancement of the visualization can be developed by collecting the information of the same data that was shared on different social networking sites.

VI. TEAM MEMBERS

The team members for the following projects are:

- Aditya Rohilla
- Anay Jain
- Harika Kondru
- Parv Sethi
- Sheen Dullu

VII. REFERENCES

[1] Shuai Chen, Sihang Li, Siming Chen, and Xiaoru Yuan. *R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media*. DOI: 10.1109/TVCG.2019.2934263. August 2019.

[2] S. Grivet, D. Auber, J.P Domenger, G. Melancon. *Bubble Tree Drawing Algorithm*. Issn: 1381-6446 in March 2006.

[3] Cheng Tang and Claire Monteleoni. *On Lloyd's algorithm: new theoretical insights for clustering in practice*. George Washington university.1280-1281.

[4] Chanhee Park, Sungjun Do, Eunjeong, LeeHanna, Jang Sungchan, Jung Hyunwoo, Han Kyungwon LeeGitViz*: An Interactive Visualization System for Analysing Development Trends in the Open-Source Software Community*. Ajou University. Pacific Visualization Symposium 2019 IEEE.

[5] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau, *Sentiment Analysis of Twitter Data*. Columbia University, April 2019.

[6] D. J. Ketchen and C. L. Shook. The application of cluster analysis in strategic management research: an analysis and critique. Strategic Management Journal, 17(6):441–458, 1996.

[7] G. Salton and M. McGill. Introduction to Modern Information Retrieval.McGraw-Hill Book Company, 1984.

[8] M. Balzer and O. Deussen. Voronoi treemaps. In Proc. IEEE Symp. Information Visualization (InfoVis), pp. 49–56, 2005. DOI: 10.1109/INFVIS .2005.1532128.

[9] Avneesh Sud, Danyel Fisher, Huai-Ping Lee. *Fast Dynamic Voronoi Treemaps*. ISBN: 978-0-7695-4112-9 DOI: 10.1109/ISVD.2010.16. June 2010

[10] http://twitter4j.org/en/index.html For scraping the data from twitter.

[11] https://dev.mysql.com/doc/

[12] https://devdocs.io/d3~4/ D3 V4 documentation