WeiboEvents: A Crowd Sourcing Weibo Visual Analytic System

Donghao Ren * Xin Zhang † Zhenhuang Wang ‡ Jing Li § Xiaoru Yuan, *Member, IEEE* ¶

Key Laboratory of Machine Perception (Ministry of Education) and School of EECS
Peking University, Peking, China, 100871

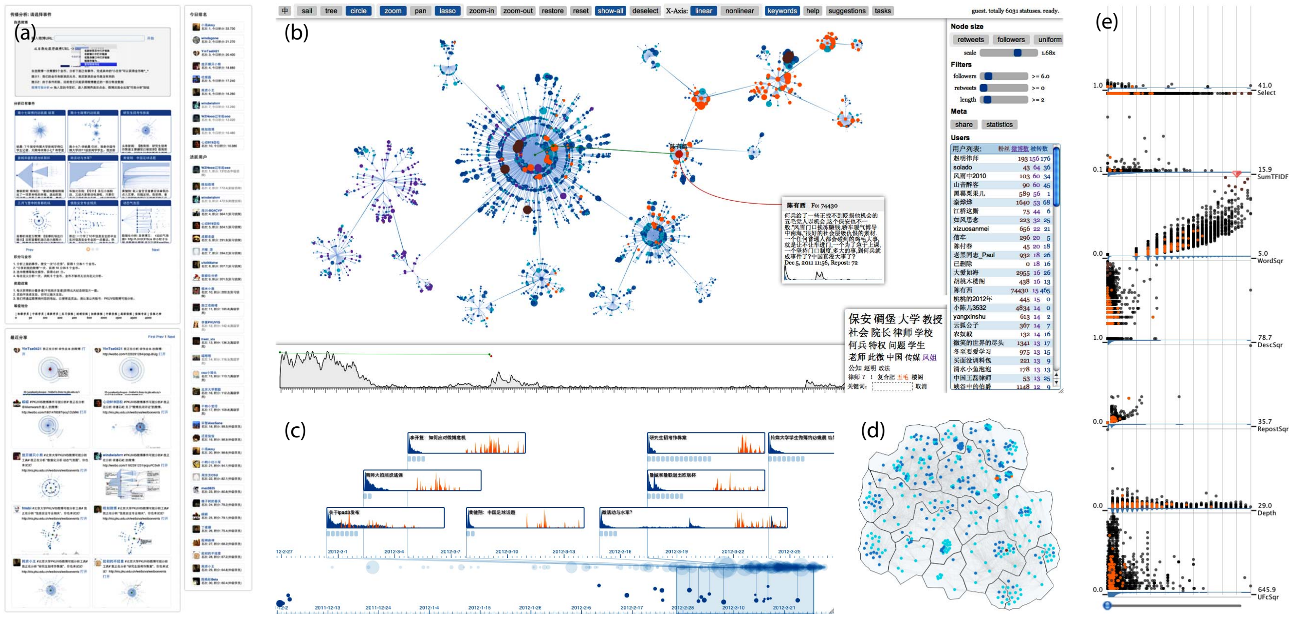

Figure 1: An illustration of the proposed Weibo visual analytic system. The components are (a) Dynamic Web Portal, (b) Retweet Tree View, used in both web-based interface and the expert analytic system, (c) Global Events Timeline, (d) Events Map, (e) High Dimensional Filters. (a) and (b) are used in the web-based interface. (b), (c), (d) and (e) are used in the expert analytic system. In this example, we're using the expert analytic system to inspect an event. Tweets with keyword "WuMao" are colored orange and tweets with higher information are colored brown.

Abstract

In this work, we propose a visual analytic system for analyzing events of Weibo, a Chinese-version microblog service. We build a system which consists of two interfaces: a web-based online visualization interface for public users and an offline expert visual analytic system which wraps the online one and provides additional analysis functions. The online interface provides an intuitive and powerful retweet tree visualization which inspires users' creativity. The expert system adopts public users' analysis results collected from the web interface, and can visualize and analyze Weibo events to a deeper extent.

Keywords: Weibo Visualization, Graph/Network Data, Time Series Data, Visualization for the Masses

Index Terms: H.4.m [Information Systems Applications]: Miscellaneous; H.5.2 [Information Interfaces and Application]: User Interface—Graphical User Interfaces; K.4.1 [Computers and Society]: Public Policy Issues

*e-mail: donghao.ren@pku.edu.cn.
†e-mail: xinzhang.pku@gmail.com
‡e-mail: wangzhenhuang@pku.edu.cn
§e-mail: eecs09lj@pku.edu.cn
¶e-mail: xiaoru.yuan@pku.edu.cn

1 Introduction

Microblogs have been very popular all around the world in just a few years. In China, Weibo, short for Sina Weibo, has been one of the most popular Chinese microblog services since launched in 2009. Compared to Twitter, retweet is much more common in Weibo [20]. Weibo users not only are more actively involved in propagating others' tweets, but also discuss others' tweets a lot by retweeting. The tree of one tweet and its descendant tweets make up a Weibo event. Weibo events contain valuable information about people's roles in the propagation of information.

Researches have built a number of visualizations for analyzing Twitter data from different domains, including general geographical situational awareness [6, 14], disaster and crisis scenarios [11, 18], and political events [9, 10]. However, none of them handles phenomenon similar to Weibo event, which is smaller in scale, but more complex in structure. To address this limitation, we present a visual analytic system for public participation and in-depth analysis of Weibo events. The system consists of two interfaces: a web-based online interface and an offline expert visual analytic system. The web-based interface allows many Weibo users to analyze both predefined events and customized events they specify in an intuitive yet powerful visualization. Apart from logging users' interactions with the visualization, we also guide users to work out mini tasks related to events they analyze. In the expert system, we visualize and analyze Weibo events with not only the original events data, but also collected users' data. It enables us to gain deeper understanding of Weibo events.

978-1-4799-2873-6/14 $31.00 © 2014 IEEE
DOI 10.1109/PacificVis.2014.38

Our work has two major contributions. Firstly we create an intuitive and powerful web-based visualization enabling general Weibo users to analyze Weibo events. We have put the system online for more than one year, and received many positive feedbacks from our users. The visualization could bring into play users' creativity and effectiveness in Weibo events analysis. It is proved by abundant visualization results from public users of the system. On the other hand, by incorporating public users' analysis data, we could better employ analytic functions of the expert analytic system. We could dig deeper into contents and structures of Weibo events more easily, and acquire some discoveries which can be difficult for automatic methods.

2 Related Work

Visual analytic (VA) systems and methods are widely used for supporting Situational Awareness (SA). MacEachren et al. [6] used both explicit and implicit geographic information and proposed a geovisual analytic system to support SA of crisis events. Thom et al. [14] applied clustering analysis and used geographical visual representation on spatial and temporal information of Twitter to conduct real-time anomaly detection. Based on real-time filtered tweets analysis. Kumar et al. [5] designed TweetTracker to gain SA in disaster and crisis from near real-time tweets trending.

Time-varying events and trends analysis and visualization with microblog data also widely exist. Marcus et al. presented a timeline-based visualization for users to browse and summarize events on Twitter [7]. Visual analytic approaches for SA [6, 14] also use temporal metadata attached to microblog text to support data filtering and aggregating. Itoh et al. [4] extract bloggers' activities and interests by analyzing blog archive, and then visualize their long-term changes in a 3D visualization. There are also works that model temporal features of conversation in microblogs [12].

Network structure plays an important role in understanding how information propagate in social networks. Ratkiewicz et al. [10, 9] applied network analysis and visualization for detecting meme diffusion patterns in election-related microblogs. Viegas et al. [15] developed history flow visualization to explore temporal dynamics of Wikipedia. Brandes et al. [1] defined the edit network of Wikipedia authors and proposed methods to better analyze their collaboration and competition relationship. In our work, we specifically visualize and analyze Weibo authors' retweet and reply structure.

There are existing studies on Chinese microblog services. For example, Yu et al. [20] analyzed temporal characters of Weibo usage, and found some differences between Weibo and Twitter. Qu et al. [8] analyzed how Chinese people use Weibo in response to a major disaster in China. Tang et al. [13] designed LifeCircle, as a summary of the long term life status of microblog users.

There are existing works on collaborative visualization. For example in [3, 17, 19, 16, 2], researchers have created systems to enable users to create, share, annotate and discuss visualization views in a collaborative way.

Most of previous work focused on summarizing and understanding microblog data from quite a large scale. It would also be interesting to understand microblog authors' micro-scale behaviors, like the scale of Weibo events. In this work, we would like to enable Weibo users to visualize and analyze Weibo events themselves. By collecting pieces of their analysis results, we could propagate their knowledge to have deeper understanding of Weibo events.

3 System Overview

Our system have two major components: a web-based online interface for public Weibo users, and an offline expert visual analytic system for in-depth analysis of Weibo events.

The online interface, based on HTML5, is intuitive and easy to use for the masses, and also inspires their creativity. After creating a visualization, the user can share it to his/hers own Weibo account.

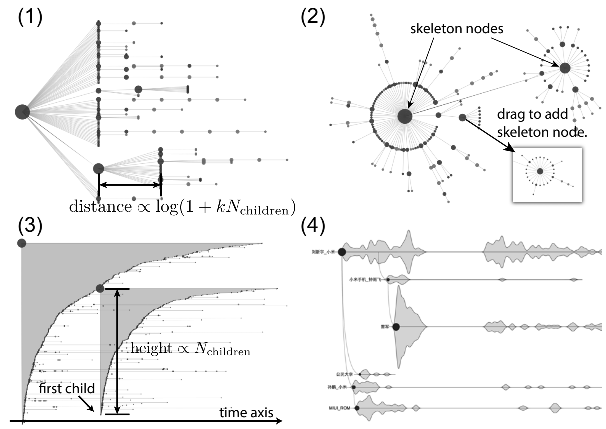

Figure 2: Three layout methods for the retweet tree view. (1) Tree layout, (2) Circular layout, (3) Sail layout. (4) Curves layout.

The online interface collects public users' analysis procedure and results, this data can be used in the expert system to support deeper understanding of Weibo events. The expert system also provides sufficient analytical functions for investigating Weibo events data.

The offline expert analytic system employs not only original Weibo events data, but also data from users of the web-based interface. Users' access logs contribute to estimation of users' interest of different events as well as importance judgement of tweets in events. With both Weibo events data and users' analysis data, the expert system provides a global events timeline which shows overview of all events and all analysis records. After loading one event with or without public user's customized parameters, the retweet tree view, high dimensional filters view, word cloud and event map in the expert system visualize one event in detail and provide linked analytic functions to enable in-depth analysis.

4 Web-based Interface

The web-based interface of our system constitute a web service for public Weibo users to analyze Weibo events. It contains an online visualization and a dynamic web portal. The web portal is the entry for the masses. It displays user guidelines, recently uploaded analysis results and a list of active users. Weibo users could log in through the web portal with their Weibo account, and then choose predefined events or input custom Weibo URLS to launch the visualization. Then the visualization will load the data specified by the user, and he/she now could explore the event visually.

4.1 Retweet Tree Layout Methods

A Weibo event is essentially a retweet tree, the retweet tree view is the central part of the web-based interface. We provide four layout methods (figure 2) for the retweet tree, users can switch between them. **(a) Tree layout**: This is a traditional horizontal tree layout. **(b) Circular layout**: In this layout, we first identify a set of important tweets by thresholding the number of retweets. The important tweets form the center of the circles, and their descendants are put around them in a circular way. The user can move the circles by dragging its central node, create new circles by selecting nodes to expand, or collapse existing circles. **(c) Sail layout**: The x coordinate of a tweet corresponds to its publish time, and the y coordinate corresponds to the index of the tweet in its parent. So the retweets of a node form a sail, the first child of the node is at the bottom of the sail, and the other children are one by one along the sail. **(d) Curves layout**: Like the circular layout, the curves layout also identify a set of important tweets, each important tweet is rendered as a timeline of its descendants. The timelines are then linked together by curves.

4.2 User Interactions

Apart from direct interaction with the layout algorithm, there are other interactions in the retweet tree view, including zooming, panning, selecting, filtering by number of retweets, highlighting specific tweets and showing their content, etc. There are also sufficient linked filtering and multiple highlighting mechanisms among all the views in the visualization, which facilitates user exploration of events. For example, selected nodes in the retweet tree will also be visualized in the timeline. Selecting multiple keywords from the keyword panel or selecting users from the user list will highlight relevant nodes in the retweet tree view in a pie chart style.

Users could easily share their visualizations with others. Annotations can be added to label their discovery or thoughts. This sharing will not only be an image with comment shared to Weibo, but also has whole settings of the result saved in our server. Later users could view previouly shared results and load them into their browser for further analysis and editing.

4.3 Crowd Sourcing

We enable crowd sourcing by logging users' interactions with the online system and providing mini tasks for users to get their opinion about the event being analyzed. The mini task is a popup box in the visualization that will popup when the visualization page is opened. There are two types of mini questions: (1) What is the role of this tweet in the event? or (2) What is the role of this author in the event? We provide 11 carefully chosen tags that users could multi-tag the tweet or author as the answer. These tags are in three categories: (1) actions with initiate, propagate, provide idea, and provide fact, (2) attitudes with opposition, neutral, and support, and (3) contents with quarrel, chat, digress, and passer-by.

5 Expert Visual Analytic System

After collecting user participatory activities from the web-based interface, we are now able to analyze Weibo events with not only original events data but also crowd intelligence from lots of users. Our expert analytic system is designed for this general goal. We devise our system for the purpose of visualizing and analyzing Weibo events data from multiple perspectives: the original event temporal feature and propagation structure, their high dimensional properties, their semantic content and public users' collective intelligence. We thus develop our system with following views: global events timeline, event graph, event map and high dimensional filter (see figure 1). The event graph view is the web-based interface embedded in the expert analytic system.

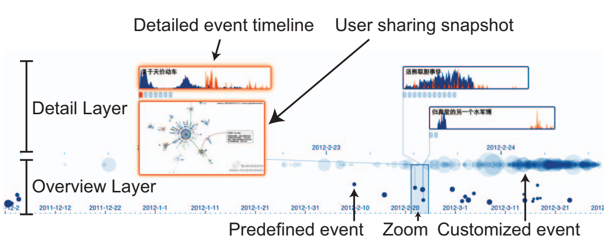

Figure 3: The global events timeline view. By hovering on one sharing label, the snapshot of that sharing pops up.

5.1 Global Events Timeline

The global events timeline (see figure 1 (c) and figure 3) provides a bird's-eye view of all predefined and customized weibo events as well as public users' analysis activities. It serves as both an overview of all events' temporal properties as well as public users' activities and the entry of going into deeper analysis.

The events timeline has two layers (figure 3): the bottom layer is an overview and the top layer is a zoomed-in view. All weibo events investigated by public users are drawn as nodes in the bottom layer. The predefined events are drawn in full opacity and customized ones are drawn in semi-transparency. Since predefined events are analyzed by many people, the y position of a node is mapped to its number of usage: the more people investigate it, the higher it will be.

5.2 High Dimensional Filters

To understand weibo events more thoroughly, we need to analyze tweets of the event through many of their properties. Our high dimensional filter view provides capability to filter tweets based on both their original data and public users' submits.

There are both structural and content properties from original tweet data. The structural properties include number of retweets, number of descendant, follower count of its author and so on. The content properties include information and strength of emotion.

With the collected users' input, we generate high dimensional role properties for tweets. For each role tag in the mini tasks, we calculate a score for each tweet. For example, for the role tag of "propagate", we could calculate the score of propagation of a tweet to be the number of submits that tag it as propagate divided by the number of submits in total.

We further aggregate role tag of "support" and "opposition" to be "attitude", which indicates support if the value is positive and opposition if the value is negative. The attitude is merged from support and opposition by subtracting them.

Due to the randomized sampling in choosing tweets and authors for mini tasks, there could be fairly amount of tweets that were never tagged by users. We just treat them as unknown and ignore them in the filters.

6 Results and Discussion

Since launched online as a public web service, our system has attracted many public users with various backgrounds. Their use of our web-base interface has shown many interesting characters.

6.1 Statistics

The online system was first opened to the public at around Feb. 2012. It was revised at Aug. 2013. Therefore there are two versions of the online system. There are no fundamental difference between the two versions, as the revision was mainly about crawling data more completely and better stability.

The usage is quiet stable. To Nov 16th, 2013, on average there are 108 sessions per day for the old version and 113 for the new version. In total there are 59652 and 7950 sessions respectively. There are 4990 shared visualization designs by 492 users in the

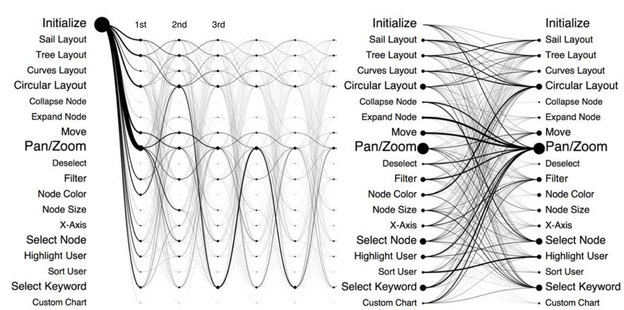

Figure 4: Left: Workflow statistics of users. Each action is represented by a node, the curves represent the interaction process, for example, we find a lot of users proceeded as "initialize, pan/zoom, circular layout, select keyword". Right: Transition graph between actions, treating it as a markov process.

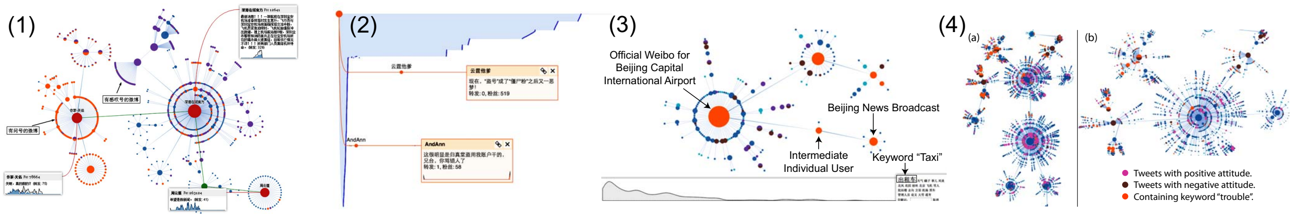

Figure 5: Illustration of results. (1) False aircraft news. (2) Robot accounts. (3) Snowing in Beijing. (4) “Student ask for interview of TV Star”

two versions. 17% of them used the sail layout, 10% used the tree layout, 71.3% used the circular layout, and 1.8% used the curve layout (which only available in the new version).

We recorded the users' interactions while they are using our system. There are 19 types actions recorded, including: layout changes, move, zoom, filter changes, selections, color mappings, axis mappings, keywords, etc. On average, there are 3 actions for each session, 8.6% of the sessions have more than 10 actions, and 0.8% of them have more than 30 actions. It is true that many users just clicked few nodes and left, but there are some professional users who used the system extensively. Figure 4 also shows the aggregrated workflow of users in the revised version. The workflow is the sequence of actions a user performed while analyzing a event. From the plot we find that users usually first pan/zoom, they also switch layout and select keywords quiet often.

6.2 Public Analysis with the Online System

People with different background use our system for quite different purposes. According to our recent survey, there are people who work on media, public opinion analysis, advertising, etc. There are also scholars using our system for their analysis. 38% of the people surveyed use the system a few times per week, 14% of them a few times per day. 13% of them many times per day. Although most users just tried the system a few times for fun and then go away, but there are professional users who use our system frequently and extensively for either their work or research.

They use our tool mostly to analyze events which have complicated structure. When they find interesting phenomena, they use our utility functions to add annotations in the visualization as labels and share them with others. For example, one user uses our tool to analyze an event talking about a news of aircraft. Using our tool, he found that while some people believe the news and feel shocked about it, others are questioning the credibility of the news (figure 5 (1)). Several minutes later, the news is conclusively belied and the event is entirely deleted in Weibo.

Another type of events commonly analyzed by such users are statuses for publicizing, especially advertisement with prize. Artificial promotion or even pure manipulation of statuses for spreading is widely visible in these events, and can be easily revealed by our tool. Figure 5 (2) shows a event with robot accounts. Some people are trying to promote a tweet by retweeting it with robot accounts. This behaviour can be easily identified by the sail layout, since the retweeting speed of the robots are generally constant, while in natural case, the speed decays over time.

6.3 In-Depth Analysis with the Expert System

With both original Weibo events data and collected users' data fed into the expert analytic system, we could analyze all Weibo events from a higher point of view as well as to deeper extent.

Figure 5 (3) illustrates an example of using the expert system to analyze one event. The event is a tweet of a piece of news about snowing in Beijing, posted by Official Weibo for Beijing Airport.

Firstly, we want to check the effects to a tweet's importantce in an event. More specifically, we want to check if there is a strong correlation between number of retweets of a tweet and its importance. We first set the node size to be proportional to retweets count. Then we use high dimensional filter view to brush out the most clicked nodes by public users to be colored yellow. Here we base the assumption that we could use selection count in the users' data to judge importance of a tweet in an event. From the visualization, we could see that two tweets of public medias (Official Weibo for Beijing Capital International Airport, and Beijing News Broadcast) are having more reposts as well as high importance based on selection count. But there is also an intermediate individual user node between the two media nodes that also attracts many selections by public users. It then falsified the proposed assumption.

Then we use filtering and brushing to color nodes with higher score of “chat” to be purple and nodes with higher score of “digress” to be brown respectively. It can be seem from the layout that chats and digresses happen mostly at close-to-leaf nodes as parts of conversation between users. In the process of brushing, the dynamic tag clouds always show keywords of the latest brushed nodes. It reveals that “Beijing” is the most occurred keyword when people are chatting, and “taxi” is the hottest topic when people digress.

Figure 5 (4) shows another example. We load the popular event “Student ask for interview of TV Star” into the system for deeper analysis. We highlight tweets with word “trouble” to be yellow, and reproduce discoveries of online users. We want to check if all the tweets mentioning “trouble” are critics of the student's action. So we brush out tweets with positive attitude (tagged to have more “support” than “opposition” by public users in mini tasks) to be pink, and brush out tweets with negative attitude to be brown. We could see from 5 (4) that the region with many negative tweets overlap a lot with areas highlighted by “trouble”. But after zooming-in and looking closer to that region (figure 5 (4.b)), we realize that they actually co-occur only in a few instances. This means that actual critics in the event are not using the word “trouble”, and the word might be used more by defenders and neutrals. By checking the content of these tweets, we finally find that “trouble” is actually used in many different context, including “there is no big trouble” or “this is the problem of the whole generation”, etc.

7 Conclusion

In this work, we have presented a new visual analytic system for visualizing and analyzing Weibo events with broader participants and deeper exploration. Our online visualization interface allows public users to conveniently visualize and analyze Weibo events. It also collects users' logs and analysis results. The offline expert system aggregates collected users' data and use them with visual analytic functions to facilitate deeper analysis of events.

In the future, we plan to support stronger analysis ability in the online interface. We could apply the system to other Chinese microblog systems. We would like to analyze deeper into the collected users' data and compare Chinese microbloggers' behaviors with existing western researches.

Acknowledgements

The authors wish to thank the anonymous reviewers for their comments. This work is supported by National Natural Science Foundation of China Project No. 61170204.

REFERENCES

[1] U. Brandes, P. Kenis, J. Lerner, and D. van Raaij. Network analysis of collaboration structure in wikipedia. In *Proceedings of the 18th international conference on World Wide Web*, pages 731–740, 2009.

[2] Y. Chen, J. Alsakran, S. Barlowe, J. Yang, and Y. Zhao. Supporting effective common ground construction in asynchronous collaborative visual analytics. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 101–110, 2011.

[3] J. Heer, F. B. Viegas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1029–1038, 2007.

[4] M. Itoh, N. Yoshinaga, M. Toyoda, and M. Kitsuregawa. Analysis and visualization of temporal changes in bloggers activities and interests. In *Proceedings of IEEE Pacific Visualization Symposium*, pages 57 – 64, 2012.

[5] S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu. TweetTracker: An analysis tool for humanitarian and disaster relief. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 2011. note: (ICWSM-11).

[6] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. SensePlace2: GeoTwitter analytics support for situational awareness. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 181 –190, 2011.

[7] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. TwitInfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 227–236, 2011.

[8] Y. Qu, C. Huang, P. Zhang, and J. Zhang. Microblogging after a major disaster in china: a case study of the 2010 yushu earthquake. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pages 25–34, 2011.

[9] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World Wide Web*, pages 249–252, 2011.

[10] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonalves, S. Patil, A. Flammini, and F. Menczer. Detecting and tracking the spread of astroturf memes in microblog streams. *CoRR*, 1011.3768 [cs.SI], 2010. http://arxiv.org/abs/1011.3768.

[11] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web*, pages 851–860, 2010.

[12] D. A. Shamma, L. Kennedy, and E. F. Churchill. Peaks and persistence: modeling the shape of microblog conversations. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pages 355–358, 2011.

[13] J. Tang, Z. Liu, M. Sun, and J. Liu. Portraying user life status from microblogging posts. *Tsinghua Science and Technology*, 18(2):182–195, 2013.

[14] D. Thom, H. Bosch, S. Koch, M. Wrner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Proceedings of IEEE Pacific Visualization Symposium*, pages 41 – 48, 2012.

[15] F. B. Viegas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 575–582, 2004.

[16] F. B. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, 2009.

[17] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[18] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 1079–1088, 2010.

[19] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 3131–3140, 2011.

[20] L. Yu, S. Asur, and B. A. Huberman. Artificial inflation: The true story of trends in sina weibo. *CoRR*, 1202.0327 [cs.CY], 2012. http://arxiv.org/abs/1202.0327.