Visualization of Social Media Flows with Interactively Identified Key Players

Xiaoru Yuan* Peking University
Zhenhuang Wang Peking University
Zipeng Liu Peking University
Cong Guo Peking University
Hongwei Ai Peking University
Donghao Ren† University of California, Santa Barbara

ABSTRACT

Microblog network, with its structural complexity and variety, prevents trivial network visual analytical tools from providing effective insight into the diffusion process of information and the effect of human participation on it. Therefore, we propose an advanced system to help users identify the key players and their various roles in the propagation of microblogs and thus analyze them based on the relations of relevant entities in between, by connecting multiple steps of filtering.

1 INTRODUCTION

In the past few years, microblogs have influenced billions of lives with all walks of people interacting via them. Its social network and user behaviour become a hot research spot.

To carry out an in-depth research on information propagation, it is necessary to understand the roles of the microblog users in the spreading process. Most of users merely act as passers-by that impose no major impact. Who arouses our real concern is a small number of users who can make a material difference, and we defined them as *key players*. Their impact can be various, such as pulling in lots of reposts, guiding public opinions, introducing information to another community, arousing criticisms and so on. Key players can reveal the backbone of information propagation. Thus, it is crucial to identify and understand them for our research.

Existing works in visualization domain which studied information propagation [3, 2] focused on temporal, spatial trends or transmission routes. As for key players in the information flow, previous reseaches just simply classified them into several categories [1, 4]. To our knowledge, there is hardly existing visualization studies deeply investigated features of key players, or that could easily differentiate various types of important users interactively.

In this work, we designed a system that integrates multiple linked views to allow a user to pick out key players through interactive filtering operation under our framework and to display their relationships and roles in the events of interest. Our system is built based on Sina Weibo, the most popular microblog service in China. The system is designed for ordinary people and thus requires no prior domain knowledge.

2 SYSTEM OVERVIEW

The objective of building the system is to help users pick out key players. The definition of "key players" is determined by users, as there is no absolute standards for it. More precisely, the system provides guidance for users to explore and discover groups of players, and each group shares common characteristics, which define the users own "key players". In our design, we put an emphasis on the flexibility, and try to cover as much features of players as possible.

The architecture of the system is shown in Fig.1, The system can be roughly divided into two parts: the back-end and the front-end. In the back-end, a crawler starts on demand to get data from Sina Weibo API and then saves them into a database. In the front-end, the four views like multi-faceted filter, task manager, entity graph and repost pathway will work together to help users to conduct exploration. They are presented in details in the following section.

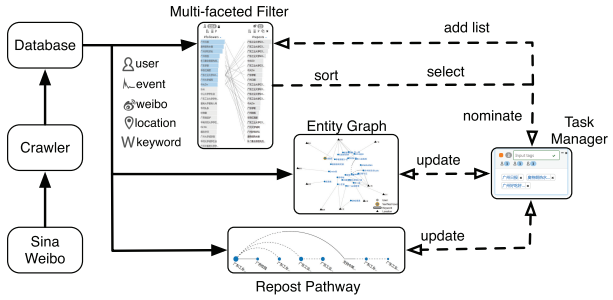

Figure 1: The pipeline of the system. The server end is in charge of crawling data and preprocessing, while the front end provides four views to support flexible filtering operation and analysis.

3 VISUAL EXPLORATION

The visual interface contains four main components, as shown in Fig.2 and Fig.3. In this section, we will introduce both the visual encodings and design considerations of each components.

3.1 Multi-faceted Filter

The most important component of the system is the multi-faceted filter, as shown in Fig.2. It accepts multiple visual operations to fulfil a specific task. Formally, a task is vector of operations, which are defined by quadruples: $(entity, sorting\ key, range, action)$.

There are many aspects of the information propagation worth exploring. We chose 5 of them: user, weibo (can be considered as a tweet in Twitter), event, location, keyword. we call them *entities*. Entities have various attributes, such as user's attributes include #followers, #followings, reposted users, etc. Entities are tightly connected with each other. For example, event is actually a series of weibos, and keyword is extracted from the content of a weibo.

In each operation, we deal with one entity type (say user), sort the entities by an attribute (say #followers), select some (say the top 10 users), and then make the next move (say see who reposted their weibos). We repeat the same process and generate the entities for the next operation. With a chain of operations, users are able to pick out the key players they expected. The filtering process can also help users form a mental impression and understanding of how

*e-mail: {xiaoru.yuan, wangzhenhuang, zipeng.liu, cong.guo, hongwei.ai}@pku.edu.cn
†e-mail: donghaoren@cs.ucsb.edu

IEEE Symposium on Visual Analytics Science and Technology 2014
November 9-14, Paris, France
978-1-4799-6227-3/14/$31.00 ©2014 IEEE

the key players are selected, and hence pattern understanding about the event involved.

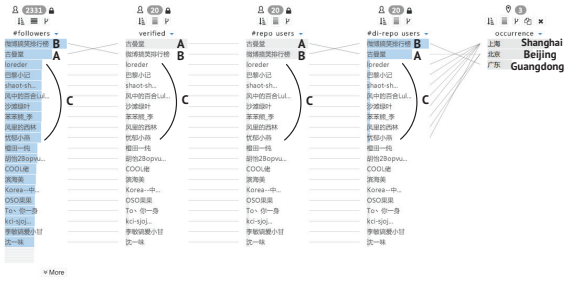

Figure 2: Multi-faceted filter. Multiple operations for one filtering task are shown.

3.2 Task Manager

There might be different types of key players, with one filter extracting one group of key players. As shown in Fig.3(b), the task manager help users manage different groups of key players resulted from different filtering tasks. Color is used to distinguish different groups in all components of our system.

3.3 Entity Graph

After potential key players are filtered, the entity graph is used to inspect relationships of relevant entities. We visualize the key players and related entities in a node-link graph. As shown in Fig.3(c), shapes differentiate types of entities; dash arcs connect entities with weak relationships, while solid arcs strong ones; colors in the nodes represent which group (filter) they belong to.

3.4 Repost Pathway

To better present how the source information spread, we visualize the route of reposts, see in Fig.3(a). We only display the positions of key players we selected from the multi-faceted filter, and their upstreams, since they are the only concerns within our system. Also, the repost pathway can verify users' theories about the key players.

The players are listed chronologically from left to right, and the arcs connect the posts with their reposts. Colors of nodes and styles of links are consistent with those in the entity graph.

4 Cases

We conducted case studies to evaluate the system. Here, we show two cases about advertising on Weibo. We succeeded in picking out several groups of key players, and had some interesting discoveries. The filtering process, along with the results, helped us understand how companies spread their advertisements.

The first case is an advertisement from the official account *A* of a company selling accessories. When sorted the users by *#reposts*, we found that *B* was particularly singled out because he took 78% of all reposts. We selected *B* as a key player who cast the great impact and owns the most #followers. Fig.2 illustrates how we use multiple operations to explore the attributes. We copied the original list and sorted the copies by different attributes. We discovered that #direct-reposts are proportional to #reposts, indicating a short chain of reposting. There are also other findings about locations and keywords distribution. The second event was originated from a dresser *D*. Noticing that apart from the source, other users are seldom reposted, we picked out the top 10 reposted players.

Comparing the repost pathways and entity graphs of both cases (Fig.3 and Fig.4), it is obvious that the process of information propagation differed a lot: the former took advantage of successive reposts from one particular well-known account, and the latter utilized many zombie accounts to popularize his ads. The key players of first case were loosely coupled, while in the second case, the source player did not have strong connections with other machine controlled accounts.

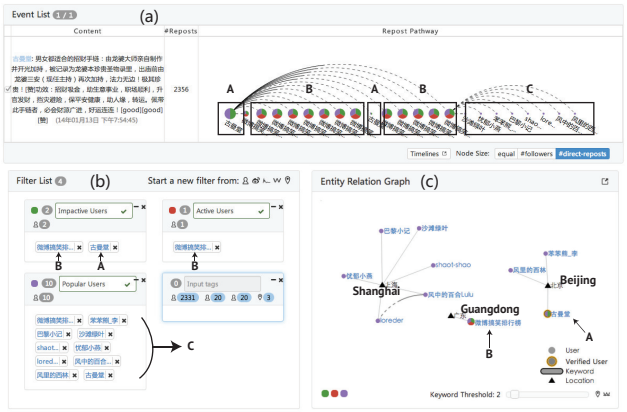

Figure 3: Case 1: Multiple filters are constructed to explore an advertisement on Weibo. Three groups of key players are picked out.

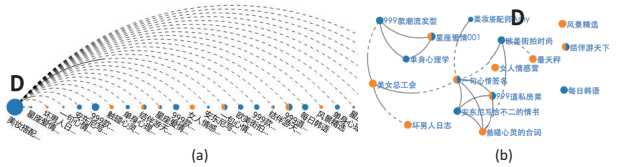

Figure 4: Case 2: Repost pathway (a) and entity graph (b) illustrate the propagation of an advertisement.

5 Conclusion

In this paper, we proposed a system for interactively identifying key players and exploring their roles in the process of information propagation. Cases show our system's effectiveness.

In the future, we will support more entity types and attributes to enhance flexibility. We will also further revise the existing framework, because there are still some features we did not cover.

Though the repost pathway can help users inspect the positions of key players in the information propagation, and find some special patterns, it still lacks a global sense of the events' propagation. We are going to create a more powerful tool for this view.

Acknowledgements

This work is supported by NSFC No. 61170204.

References

[1] M. Cha, F. Benevenuto, H. Haddadi, and K. Gummadi. The world of connections and information flow in twitter. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 42(4):991–998, 2012.

[2] R. Donghao, Z. Xin, W. Zhenhuang, L. Jing, and Y. Xiaoru. Weiboevents: A crowd sourcing weibo visual analytic system. In *Proceedings of IEEE Pacific Visualization Symposium (PacificVis 2014)*, 2014.

[3] Q. Li, H. Qu, L. Chen, R. Wang, J. Yong, and D. Si. Visual analysis of retweeting propagation network in a microblogging platform. In *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction*, pages 44–53. ACM, 2013.

[4] P. Xu, Y. Wu, E. Wei, T.-Q. Peng, S. Liu, J. J. Zhu, and H. Qu. Visual analysis of topic competition on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2012–2021, 2013.