Visualization of Social Media Flows with Interactively Identified Key Players

Xiaoru Yuan, Zhenhuang Wang, Zipeng Liu, Lijing Lin, Cong Guo, Siming Chen, and Donghao Ren

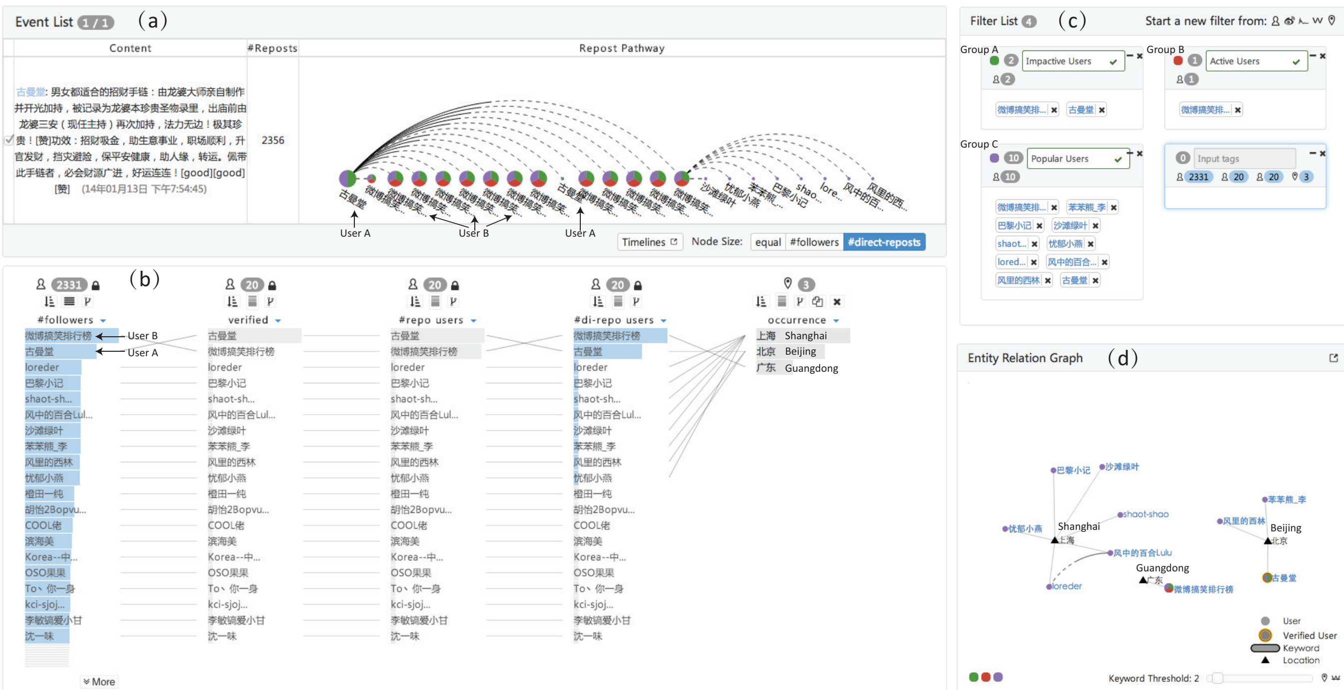

Fig. 1. The main interface of our system, the components are (a) event list and repost pathway, (b) multi-faceted filter, (c) filter task management panel, (d) entity graph.

Abstract— As microblog network has complex and various structures, trivial network visual analytical tools can not provide much insights of how information diffuse and how people influence the process. We propose a visual analytical system which allows users to interactively identify key players who play various roles in the propagation of microblogs. We first conducted an investigation on features of key players, and then built up a multi-faceted filter interface to guide the exploration. By connecting different steps of filtering, users are allowed to explore key players according to their features gradually. Meanwhile, leaded by candidate key players, users can analyze the propagation of microblogs by exploring the relations of relevant entities. We finally show how it works and evaluate our system with three real cases.

Index Terms—Microblog, Information Flow, Key Player, Multi-faceted Filter

1 Introduction

In recent years, microblog has gained worldwide popularity rapidly. Researchers have shown many ways to explore microblog as it has a hug amount of invaluable data, from which both physical world and human society can be perceived. But there are still much more potentials we could dig from microblog.

Twitter, the most popular microblog service, has made huge differences in many scenarios, playing an increasingly important role in human's life and transforming the ways we communicate gradually. There are countless examples to verify the influence. The news of Osama bin Laden's dead was first broken on Twitter; Twitter accelerated the development of Egyptian revolution; during Japan Earthquake in 2011, Twitter became an invaluable tool, and numerous people came to Twitter to keep in touch with relatives as well as get latest information; even in the 2012 US presidential election, Twitter was employed to bridge the distance between voters and candidates. There are still lots of events underscoring the growing influence of Twitter. In China, Sina Weibo, one of the most popular microblog services, also has a profound effect on people's daily life and the society.

On a typical microblog platform, millions of users form a huge network which is the backbone structure of information flow. Information on microblog always spreads along edges of the network, from one user to another user. The action of reposting (also called "retweet" in Twitter) means to forward the information to your followers, which is the most common way to spread information from the original author

- *Xiaoru Yuan, Zhenhuang Wang, Zipeng Liu, Lijing Lin, Cong Guo, and Siming Chen are with School of EECS, Peking University. E-mail: xiaoru.yuan, wangzhenhuang, zipeng.liu, lijinglin, cong.guo, siming.chen@pku.edu.cn*
- *Donghao Ren is with University of California, Santa Barbara. E-mail: donghaoren@cs.ucsb.edu.*

to those who do not know him / her.

To carry out a more in-depth research on information propagation, it is necessary to understand users' roles in the spreading process. However, most of them act as passers-by indeed as they haven't impact the event's development greatly. Therefore, our real concern is a small number of users who have made a great difference, and we call them *key players*. Their impact can be various, such as bringing in lots of reposts, guiding public opinions, introducing information from a community to another and so on. With key players, analysts can easily access the backbone of information propagation process. Therefore, identifying and understanding key players is important in the research of microblog broadcast.

Existing works in visualization domain studied on information propagation [6, 15] focused on the temporal, spatial trends or transmission routes. As for key players in the information flow, some works just simply classified them into several categories [3, 30]. To our knowledge, there are few of studies which have deeply researched features of potential key players, or could support users to pick out various types of important users interactively.

In this work, we conducted an online survey to collect potential features of key players, after which we summarized the features provided by crowds. However, most were stated in natural language which were too flexible for a deterministic filtering within the competence of existing algorithms. Therefore, we proposed a formal definition, covering a considerable part of features. We then designed a system that integrates multiple linked views to allow users to pick out key players through interactive filter operation under our framework. The entity graph, repost pathway and repost timeline enhance the ability of analysing concrete roles which key players have played. With the help of our system, analysts are able to conduct an effective exploration of key players, to have a better understanding of the information flow.

Our major contributions are:

- We presented the potential features of key players summarized from an online survey, and described them with a defined formal format.

- We proposed a system consisting of multiple linked views that allows users to interactively conduct filter operations to pick out key players in the process of information propagation, and explore the relationship of users in depth.

- We showed three real cases to explore key players in information propagation using the system.

In the following sections, we start by reviewing related work. Then we introduce our method of key player identification and our system. After presenting cases, we found with the system, we conclude the paper with a discussion on the pros and cons of our approach, as well as the future work.

2 Related Works

Existing works related to us can be mainly classified into three categories: visualization of microblog, information propagation on microblog and user roles on microblog. This section will briefly discuss related literatures.

2.1 Microblog

After launched in 2005, Twitter is proliferating in a tremendously fast speed and has shown huge potential of microblog to the world in just few years. It is now part of the indivisible digital world where people live in everyday. Its increasing influence attracts lots of researchers from different domains to study microblog. Their works have further proved the inestimable value contained in the huge amount of short messages people post everyday on microblog. For examples, microblog data can be used for disaster and crisis detection [23]. Lotan et al. [16] showed how microblog can influence people's social and political life at the level of a nation. Hu et al. [12] claim that twitter spreads news faster than traditional media and has potential as a new media. As a consequence, due to nature of real-time and widely distributed users, visual analytics methods for microblog data is good at Situational Awareness (SA) [13, 17, 26].

Time-varying is one of the main features of microblog data. Kate Starbird et al. [24] employed timeline to examine Twitter activity during 2009 Red River Floods. TwitInfo [18] tracked time-varying trends of real-time Twitter information flow, looking for the peak of events and make labeling through meaningful text from the content of tweets.

Following links people on microblog, and constructs a tremendous network. Information flows along the edge of network, thus exploring network structure can help us understand how information diffuse on microblog. Node-link diagram is a common method to visualize network structure [8, 20]. In addition, adjacent matrix could also serve to reveal the structure [10].

Many techniques have been incorporated by researches to understand content of microblog data, including key words extraction and filtering [20, 23], events extraction [18, 23] and sentiment analysis [18]. There are also more advanced visual analytic techniques, like text clustering and particle-based visualization [1] or dynamic word cloud visualization [5], that are used to analyze large and dynamic text corpus. Introducing text mining techniques, especially topic modelling, visualization techniques can reveals more meaningful information and explorer the evolution of topics [4, 7, 25].

In China, microblogging services similar to Twitter are also flourishing rapidly. Weibo is the Chinese word of "microblog". Sina Weibo is one of the most popular Chinese microblog services, which launched in 2009. It provides comparable functions to Twitter, centering around short public messages. Users can post messages up to 140 characters, in which "@" can be used to mention other people, and "#" starts a hashtag. Repost is similar to retweet in Twitter, it is the main approach to spread information to others. One weibo and all its descendant reposts make up a tree, also called Weibo event. As for relationship, there are two concepts: followers and following. When you follow someone, you are subscribing to their weibos in your news feed. You will be shown in his/her followers list, if he/she is in your following list.

2.2 Information Propagation on Microblog

Information spreads through friends and followers in the social network. Information propagation in social network is studied for decades, especially with the rapid rise of microblog, researchers from multiple domains are attracted. Scientists study the information flow from different perspectives like model, prediction, and mechanism [11, 22, 31].

Kwak et al. [14] suggested that the information flow followed another pattern that different from the traditional social network. Therefore, exploring the pattern of information flow on microblog would be interesting and meaningful.

Some existing works provided visualization methods to reveal the process of information diffusion in an intuitive way. With the hybrid metaphors of node-link and circular treemap, Google+ Ripple [28] visualized the information flow on Google+. Some other novel and designed metaphor, like sunflower [2], flow [29, 32] to traced information propagation in Twitter. Based on the structure of the social network, Li et al. [15] revealed dynamic attributes of information propagation. Using another novel and special designed metaphor, whisper, Cao et al. [2] traced the information propagation from three aspects: the temporal trend, social-spatial trend, and community response. WeiboEvents [6] provided several views to visualize the retweeting tree of a source tweet, indicating the process of propagation of an event on Weibo platform.

2.3 User Roles on Microblog

Several works studied on individual users' or communities' impact on information flow [21, 27]. Further more than individual users, some studies classified users' role played in the propagation. Cha et al. [3] grouped influential users into three categories: mass media, who had ability to reach a large number of audience; grassroots, who were spread throughout the network but have not so much followers; and evangelists, like opinion leaders, who actively participated in information diffusion. Analogously, Xu et al. [30] used media, grassroots, political figures in their work following the opinion of domain experts.

Although existing works provided various classification methodologies and characteristic of them, they only take limited factors into consideration and make simple classifications. In fact, the role of users play in social network is much complicated. Our work takes multiple attributes of users into full consideration, summing up the features of key players from an online survey. Further more, we provide a system to interactively extract potential key players with flexible filtering.

3 System Overview

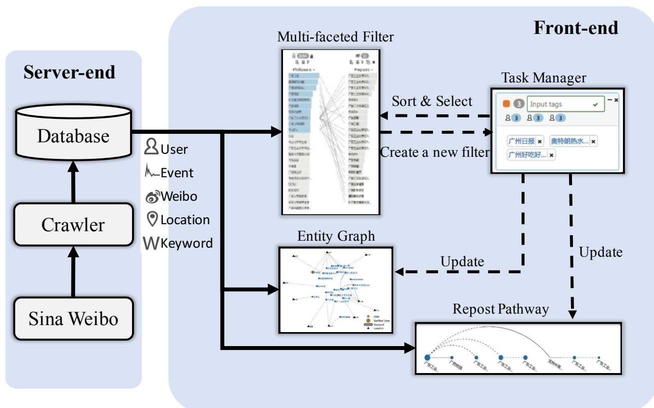

Fig. 2. The pipeline of the system. The server end is in charge of crawling data and preprocessing, and the front end provides four views to support flexible filtering operation and analyzation.

Figure 2 is the pipeline of our system. The system can be roughly divided into back end and front end.

According to the requirements of data, a crawler is started on demand to get data from Sina Weibo API. It will fetch all reposts of a specific source weibo, personal information of users involved in the events, as well as their following relationships. All data crawled is stored in our database.

On triggering a crawling task, a new dataset is established to hold continuously coming data. In the server end, we generate necessary indices over the data based on pre-defined queries. Meanwhile, all tasks, such as extracting keyword and processing the newly coming pieces of weibo, run automatically and incrementally, preparing well-organized data for future use.

In the front end, there are four views after data loading: multi-faceted filter along with a task management panel, entity graph, and repost pathway. The user can first pick out some potential key players who share some common features step by step using the faceted filter, and inspect their relationships in the entity graph and repost pathway to generate a mental image of the candidates. After that, he/she can modify the previous selection, tag the candidates, or start a new filter to explore other features in the same way.

In the end, several groups of players are dug out, with each group targeting one feature. In other words, the user picks out some categories of key players in an event, with a deeper understanding of how information was propagated in a twitter-like social network.

4 Key Players Identification

Information propagates on microblogs via retweet, which spreads tweets from one user to others. Therefore, through a series of retweet, a user could access to information which was posted beyond his followers. In this social network, users dominate the process of information diffusion, but they usually play different roles. Therefore, in order to identify key players, it is necessary to understand their features in various situations.

4.1 Online Survey

As there are a thousand Hamlet in a thousand people's eyes, this question will not have final conclusion. We try to discover common features that most people agree on. To better define features of key players, we carried out a questionnaire survey on the Internet to collect features of key players, and meanwhile ask respondents to evaluate features randomly selected from others' answers as a means of verification.

During a week the survey opened to public, 350 questionnaires were collected. We list features with high average scores and classify them in several categories as follows.

1. **The nature of accounts**
   1.1 Accounts of official organizations, like governments and companies.
   1.2 Accounts of celebrities.
   1.3 Verified accounts, who registered with real identities.
   1.4 Active users who posted many weibos.
2. **Position in community**
   2.1 Users with a lot of followers. These users are able to influence the mass.
   2.2 Users who connect two communities acting as a bridge. They have the potential to spread information from one community to the other.
3. **Text Semantics**
   3.1 Users who propose different opinions, they are possible to lead another topics or change the standpoint of followers.
   3.2 Users who are tagged (@) by others many times in an event. They are closely related to this event.
   3.3 Users who claim to be witnesses or have experiences of the event, especially accidents.
4. **Information propagation**
   4.1 Users who bring many reposts.
   4.2 Users who bring information from a community to another one and cause a number of retweets. They are often labelled as bridge users.

4.2 Filter Operation Description

Results of the survey gave us a large amount of potential features, while most of them were defined in natural language.

Therefore, it is necessary to translate features described by words to formal definitions. Only in this way can the implementation have deterministic behaviour.

Entity An entity represents an object to be observed and explored. Currently, we support 5 different types: user, event, weibo, keyword, location. Each entity contains several attributes. All the entities and their attributes are listed in Table 1.

These five entities are commonly referred to when we are looking for interesting players in an event. They have their own attributes, and meanwhile, they are tightly connected with each other. For example, a weibo is posted by a user; an event contains weibos with post-repost relation; keywords are extracted from text of weibo. Deep understanding of every entity type and relationships between them is crucial to our solution.

Operation Operation is one step in the filter process. To describe every facets of an operation in a standard format, we put all information the system needed to work with into a quadruple: $(entity, sortingkey, distribution, action)$. $Entity$ is one of five entity types we currently support; $sortingkey$ indicates what attribute of the entity we are interested in; $distribution$ represents places of selected players in the list; $action$ represents how we expand the selected to a new list.

For example, $(user, \#followers, [0-3\%], listfollowers)$ is a valid operation. It means sort users according to the number of their followers, then choose the top 3%, and add a list of their followers.

Task A task is a series of operations to nominate a group of key players, which can be represented as a vector of operation quadruples. Labels can be attached to a task to indicate the common features of the group. In Table 2, we list some examples of filter tasks.

5 Interface

Based on the questionnaire results, we are aware of a considerable amount of features of key players. A user of our interface should be able to pick out a specific group of players of the same type, as well as to

Entity	Attributes	Actions
User	**#followers**, the number of followers.	a1. select as candidates a2. copy to a new list a3. list followers a4. list locations a5. list events the users participated in a6. list weibos the users posted
	#following, the number of followings.	
	#following / #followers, the ratio of following to followers, it is an statistics that reveals the type of users. If the ratio is large, the user could be a spam accounts, instead the user could have influence.	
	#mutual-following, the number of mutual followings.	
	verified, the account is verified or not.	
	#events-involved, the number of events the user involved.	
	#weibos, the number of weibos the user posted.	
	#reposted users, the number of users who repost weibos that posted by selected users.	
	#direct-reposted users, direct-repost is a subset of repost, it refers to reposts that directly forwarded the target weibo instead of reposting from other's forward.	
	#direct-reposted users / #reposted users, the ratio of direct-reposted users to reposted users. If the ratio is much less than 1, the user may perform as bridges in information propagation.	
Event	**#tweets**, the number of weibos evolved in the event.	b1. list all weibos in the events b2. list all participants in the events b3. list all keywords in the events
	#participants, the number of users who evolved in the event.	
Weibo	**time**, the post time of the weibo.	c1. list all reposts of the weibos. c2. list all direct reposts of the weibos. c3. list authors of the weibos. c4. list keywords of the weibos. c5. list events the weibos belong to.
	#comments, the number of comments of the weibo.	
	#likes, the number of users who like this weibo.	
	text length, the length of text of the weibo.	
	#direct-reposts, the number of direct-reposts.	
	#reposts, the number of reposts.	
	#direct-reposts / #reposts, the ratio of direct reposts to reposts.	
Keyword	**frequency**, the frequency of the keyword.	d1. list relevant weibos
Location	**frequency**, the frequency of the location.	e1. list relevant users

Table 1. Entities, their attributes and enabled actions.

Feature	Task Vector	Label
A lot of followers	$<(user, \#followers, [0-x\%], a1)>$	Popular user
A lot of weibos	$<(user, \#weibos, [0-x\%], a1)>$	Active user
Verified	$<(user, verified, true, a1)>$	Verified user
#followings $\gg$ #followers	$<(user, \#followings/\#followers, [0-x\%], a1>$	Potienal zombie user
Involved in many events	$<user, \#events-involved, [0-x\%], a1)>$	Active participants
Users directly reposted A	$<(user, *, A, a6), (weibo, *, *, c2), (weibo, *, *, c3)>$	Disseminator of A
Reposted by influential users	$<(weibo, \#direct-repost/\#reposted, [0-x\%], c3), (user, *, *, a1)>$	Information bridge

Table 2. Examples of filter operations. "*" means any, not determined.

manage and compare various types of players using multi-faceted filter and task management. Then, relationships of corresponding entities and post-repost relation of candidates are presented, in order to help the user create a mental image of how the events started, fermented, and spread.

Our design mostly aims to support both task management and a collection of filter operations.

5.1 Multi-faceted Filter

In this part, we have two major considerations:

1. Explore entities with a specific attribute.
2. Connect multiple attributes of entities and different entities.

For the first one, ranking list is a simple but powerful visualization for presenting entities with properties, especially for numeric data. Usually, a user mainly cares about the rankings or relative relations of attributes, and thus we show a bar with its length indicating the attribute value of each entity.

Selected entities are highlighted by brighter color, and are later used for expansion. To connect different operations so as to form a chain, lines are drawn between two relevant entities: same entities with different attributes or different entities with certain relation.

5.1.1 Between Lists

When connecting two adjacent lists of the same entities with different attributes, multiple facets of the entities are displayed. For example, comparing the number of followers and the number of reposts of all users, we might discover that these two properties have a positive correlation since users with more followers have more chances of interactions. When connecting two lists of different entity types, the user can grasp the relationships between them so that he can extend the filter process for further exploration. In the case of user and weibo, the user and the weibos he posted are linked, so attributes of the weibos can be analyzed, such as selecting some interesting weibos and expanding their reposts. Furthermore, the user can list the authors of these reposts. From user to weibo, and back to user, we know who reposted the weibos.

5.1.2 Within A List

While connections between neighboring lists are to keep on the course of filter, helpful functions for one list are to extend the ability of the ranking. In the normal situation, users are more interested in entities with attributes of higher values (e.g users with #followers greater than 10000), but sometimes paying attention to lower ones might result in surprises (e.g outliers). Thus both descending and ascending sort is supported. To increase scalability of lists containing thousands of

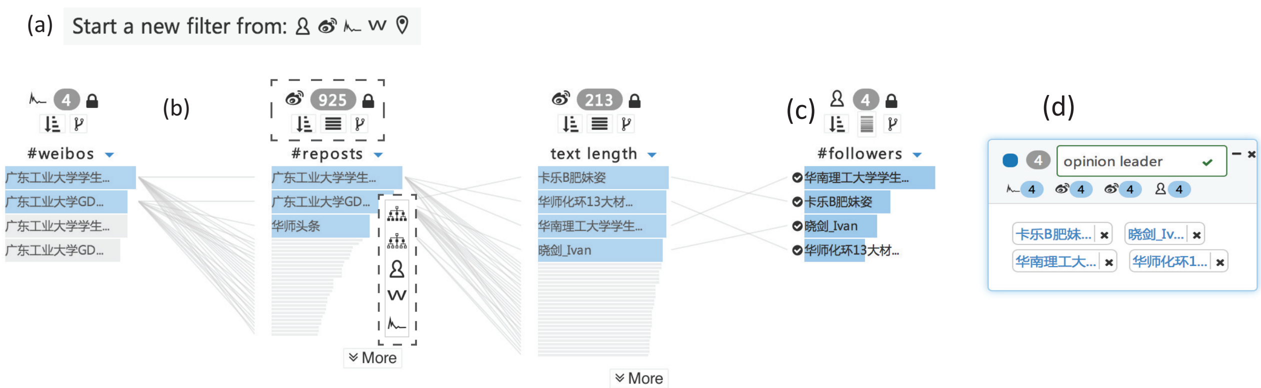

Fig. 3. Multi-faceted filter interface. (a) **Start point**: the user can crate a new filter beginning with a list of any entity type. (b) **Filter expedition**: select entities, explore their attributes, and append connected lists. There is a list header and control buttons (sort, toggle table lens, fork) specified in 5.1.2. A lock icon indicates prohibitions on the list (see 5.1.3). One action menu is highlighted in the black box. (c) **Destination**: nominate users as candidate key players. (d) **Filter panel**: a summary of the filter process and candidates are shown, and the user can input a tag to indicate a common feature of corresponding candidates.

entities, users can employ "table lens" [19] technique to shrink rows, and extend the number of rows or collapse some. List duplication comes to help when comparing different attributes of the same entities with the help of connections between. Also, sharing part of the filter process is enabled by branching (forking), which copies all the process up to the operating list to a new filter where user can continue picking out another group of key players, analogous to create a new branch in Git. Finally, a list can be removed in case of misoperation.

5.1.3 Rules

Operations are strictly carried out from left to right like a directed information flow so that a user can enjoy a mental map of the whole filter process. To prevent the filter direction from going backwards, we also impose prohibitions on them. It is not allowed to remove a list in the middle but only on the rightmost, nor to re-select entities in previous lists. A list is appended-only, which means it directly relates to its left but has nothing to do with a distant one.

5.1.4 Multiple Selection Mode

One critical part of the operation is selecting entities in a list in a flexible manner. Thus we enable multiple selection mode. Like the interaction method in Microsoft Excel, the user can select one entity by clicking, or select a number of successive entities by dragging an area. With the "shift" key held on keyboard, discrete entities can be selected, and with the "alt" key held, selecting entities are kicked out. Since the final result of a filter process highly depends on careful selection in each operation, we try to provide the ability to ensure the possibilities of all combinations of selection.

All these functionalities and limitations are aimed to make sure that user conduct the filtering step by step, and finally reach the end, just like an expedition.

5.2 Task Management Panel

Organizing filters requires an overview which can capture their most important features.

In a typical case, there are usually different types of key players. Each task is dedicated for one target type of players, which is presented as a block in the filter list (also task list). The header (See d in 3) of a block includes the summary of the filter process, consisting of entity type and number of selected entities in every operation. The body of a block holds all candidates of key players, where we can see detailed information of players and single player removal is supported. Another important feature of a filter (task) is the tag of candidates. The user can input a tag to indicate the common characteristics of the candidates, which is later sent back to our task database. One important future work is to recommend tags to the user once we collect enough solid tags.

5.3 Entity Graph

After the user filtered the potential key players, we need a view to inspect relationships of all relevant entities. Although previous works, such as [9], have presented wonderful tools to analyze people's relationships, but here they do not suffice due to multiple types of entities, which have inner connections with each other. We visualize the key players and their attributes (entities) in an entity graph. There are three types of node in the entity graph: player, keyword and location, with other information, like user verified, added to support the tasks of identifying key players.

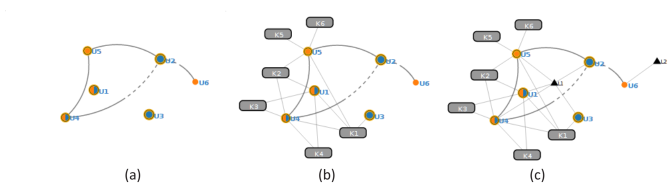

Fig. 4. Entity Graph. (a) An initial layout. Dash line means weak relation and solid means strong relation. In the graph, U4 and U5 are fan friends of each other, and U4 is fan friend of U2 but U2 is not of U4. And only U6 is not verified user. (b) Add keyword attribute. Rectangle node represents keyword. (c) Add location attribute. Triangle node represents location. Only U6 is located in L2, others are in L1.

Force-directed layout algorithm is applied to layout the entity graph. We set the user node's weight proportional to its followers and use a constant value for keyword node and location node. Weight of player-player link is set by the reciprocal of the follower's friend number, and weight of player-keyword link is set by keyword occurrence times in the player's tweet content. After set all the nodes' and links' weight, we first run force-directed layout between player's nodes and after all player nodes become stable, a further layout used for keyword and location nodes. Whenever user delete or add a new player, the layout process is calculated again.

Shape is used to differentiates different types of nodes, like circle for user, rectangle for keyword and triangle for location. Then we use different colors to distinguish different groups, if one player belongs to multiple groups, we divide the circle to pie chart with coordinate group colors. Other information which is useful in key players identifying are also encoded, such as orange stroke for verified player. We encode links in two kinds: player-player link and player-attribute link. Player-attribute link is directly plotted by a straight line. And for the player-player link, we use a slightly darker arc links two players that have relation, and a dash arc is used to represent a relatively weak relationship, solid arc represent strong relationship as shows in Figure 4(a). All the links length is proportional to the nodes' correlations.

In the entity graph, we support several interaction operations for user's exploration. User can hover or click on a node to view and select it. For convenient use, we also support user to filter players by groups or filter node by types. And for semantic finding, user can filter keywords by the occurrence times.

5.4 Repost Pathway

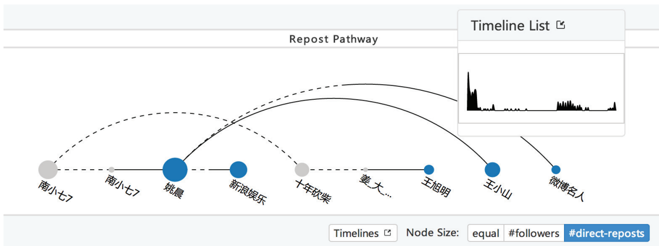

Fig. 5. Repost pathway presenting the post-repost trails of relevant users.

After discovering potential key players with the interface described above, the user is able to inspect the places they stand in the weibo information propagation, as a validation of his decision in selecting key players.

As shown in Fig 5, we update a view presenting post-repost relations each time the user add or remove candidates. Weibos posted by candidates and relevant users are lined up in the row of each event chronologically. Then links are drawn between them: straight lines for neighboring ones and arcs for others. Dotted and full lines indicate follower-following relationships of user pairs, in a manner same as that of the entity graph. The node size is encoded by the number of followers, the number of the weibo's reposts, or neither (in the case of equal sized nodes), and color is consistent with all other views. The path of a weibo from its source down to reposts is highlighted on hovering, and the trend of its following reposts is popped out on clicked.

6 Implementation

Our system is built in browser-server architecture using RESTful web services. Data is stored in MongoDB, a kind of NoSQL database that stores data in JSON-like documents. The server is in charge of data crawling and data preprocessing, as well as performing queries in the database, refines the data structure to general JSON format and transfers it to the browser interface. In the client side, we built a web-based user interface implemented with HTML and JavaScript. By doing so, we allow public access to our system with various types of web browsers.

7 Cases

To evaluate our analytical system, we chose three real cases in Sina Weibo which behaved fundamentally differently. We succeeded in picking out some key players with various characteristics, as well as interesting discoveries. Both the results and process of filtering helped us understand how these events started and propagated.

7.1 Student Asking for interview of TV Star

In Sina Weibo, there was an interesting event a student asked for an interview of a TV star, originated from a weibo of the student. In the weibo, she mentioned a famous female TV star in China and asked for an interview with her to finish her homework. This weibo was not noticed by the TV star until 20 days later when the original tweet has been retweeted for more than 600 times. When the tweet was finally noticed by the TV star and the request for interview was accepted in her repost, a burst occurs of the diffusion and the repost brought more than 10000 retweets in just one night.

We analyzed the event in our system. First, we established a filter to pick out most popular peoples, following the task vector of $<(user, \#followers, top3, a1)>$. The TV start (user B) and two media accounts (User C and User D) were selected out. A burst after the TV star reposted the original weibo could be observed from the timeline. Meanwhile, media accounts also made small peaks.

Next, we set up another filter to find out who brought most reposts. The vector $<(user, \#reposted - users, top3, a1)>$ was adopted here. The student (User A) and the TV star were filtered out with no doubt. Another selected user (User E) had many interesting features to be worth noticing. The reposting path indicated that he not only earned lots of reposts but also took part in this event many times. If we further observed the entity graph, it was easy to found out that keywords he used were different from other users, as shown in Figure 6(b). Reaching into the content he posted, we noticed he was criticizing the student for not being polite and respectful. With this phenomenon observed, a conclusion comes that he was an opinion leader in this events for he raised negative voice, and earned the spread of his opinion by active participation.

7.2 Advertising on Microblog

The rapid growing number of users makes microblog an attractive platform for advertising. Compared with traditional advertising platforms (e.g. TV, newspaper, radio), the propagation pattern on microblog is different.

In this case, we provided two events to illustrate how to use our system to explore the propagation of advertisement on Sina Weibo.

The first advertisement were originated by the user "Gu Man Tang" (User A), see Figure 1. We started a filter to observe users reposted this weibo. When users were sorted by *#reposted users*, we saw that the event had 2331 reposts in total (it equalled to the reposts of original weibo), while the number of reposts of the second user (User B) was 1798, 78% of all reposts. Hence we picked out the two users with the vector $<(user, \#reposted - users, top2, a1)>$. Next, when we sorted users by the key *#weibos*, the most active user was User B who reposted the weibo 13 times, while others only appeared in this event two times at most. Therefore, we made the second filter following the vector $<(user, \#weibos, top1, a1)>$ to select out User B. Furthermore, we wanted to explore users with lots of followers. We used vector $<user, \#followers, top10, a1>$ to filter out these users. Apart from the simple filter that only adopted one operation, step-by-step operations would help analysts better understand key players. See the Figure 1(b), we sorted users by the number of followers and then expanded them to a new list. In the second list, it was easy to notice only User A was verified, while the User B was not verified though he had more followers than User A. Observing the third list sorted by the number of reposted users and the fourth list sorted by the number of direct-reposted users. It was obvious that the number of directed-reposted users was proportional to the number of reposted users except the top two users (User A and User B). It suggested reposts were mainly directed-reposts, not many undirected-reposts. The fifth list indicated most of users came from Shanghai.

After the three groups picked out, we turned attention to repost pathway. User B reposted a lot and he always forwarded from the original weibo every time he reposted. As for other popular users, they all forwarded the last reposted weibo of User B. However, these popular users didn't make much influence to the propagation. In addition, they were neither User B's followers nor following. Figure 1 also revealed users in Group C hardly followed each other. When reading the content, we found text of all the 13 reposts published by User B were the same. The entity graph also revealed a special phenomenon that most of users in Group C were in Shanghai, but User A and User B were in Beijing and Guangdong province respectively.

Another event was original posted by a dresser (User C). Likewise, we first used vector $<(user, \#reposted - users, top10, a1)>$ to single out the user who produced the most reposts. But unlike the former event, apart from the User C, others users in the top 10 just had a few reposts, making little influence. We also built the second filter with the vector $<(user, \#followers, top10, a1)>$ to select out most popular users. See Figure 7(a), from the repost pathway, we found all the reposts were forwarded directly from the original weibo. Almost all of the these transmitter had no relationship with User C, but connections among these users were much more close (see Figure 7(b)) than the

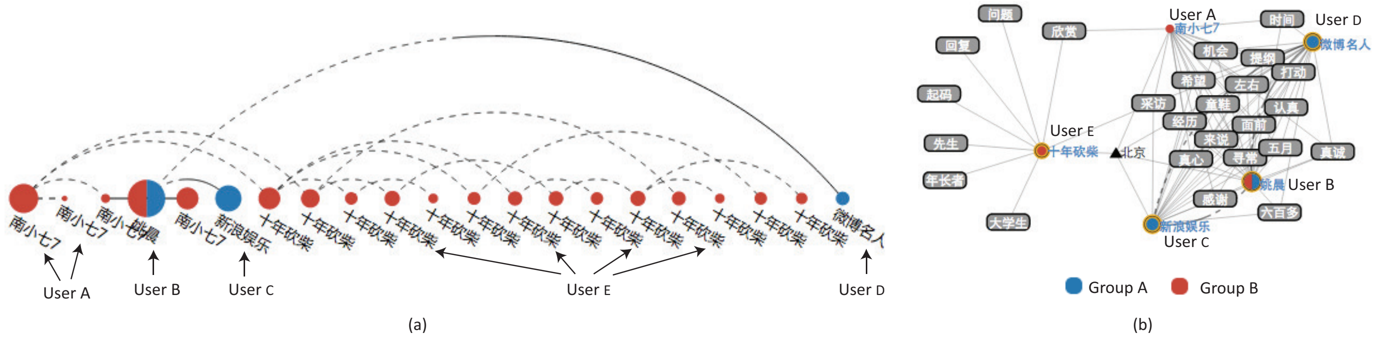

Fig. 6. Repost pathway (a) and entity graph (b) of the event "student asking for interview of TV star".

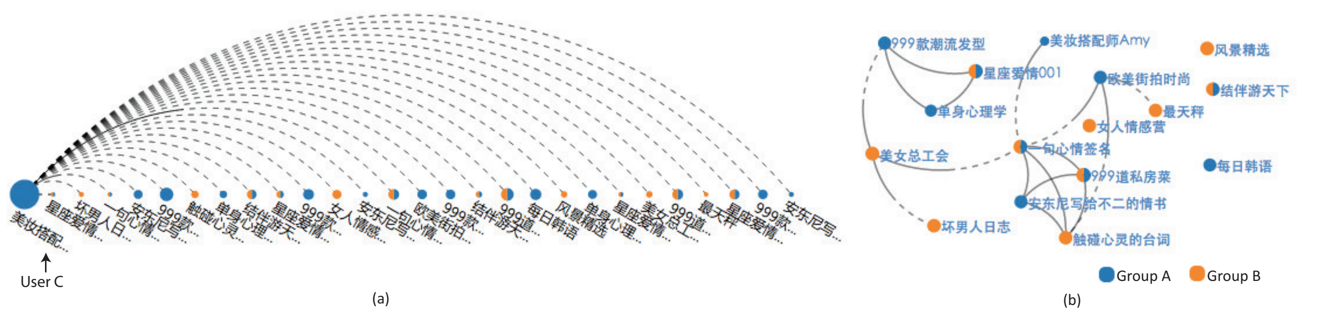

Fig. 7. Repost pathway (a) and entity graph (b) of the event of an advertisement originated by a dresser.

above event.

The two events about advertising on microblog showed different patterns of propagation. The former one relied on lots of accounts special for marketing, while the latter one relied on a popular account. This case proved our system were able to identify different type of key players and distinguished different propagation patterns.

7.3 Hot Water Supply Insufficient in Freezing Days

This case is about hot water supply problem in Guangzhou Higher Education Mega Center in the coldest days. Students on campus complained about the heating company and natural gas supplier on Sina Weibo, which soon created a heated trend, pulling in official student organizations and also newspapers.

We crawled relevant six events as our analysis samples. Firstly, we sorted all the users by the number of reposts to have a glance at who were the most influential ones. We selected a dozen of the highest ones and listed out their locations. We could see from Figure 8 that Guangdong dominated the user locations, which indicated that the events happened in Guangdong Province. Moreover, when we listed all user locations, we found that there were people from outside of Guangdong participated in it. Picking out Beijing for further exploration, we discovered little number of reposts of unverified users from Beijing, such as user A, two thousand kilometers away from Guangzhou. With these two findings, we could conclude that this event was highly related to one geographical region.

Next, we started a new filter to see what was interesting in people with great number of followers. As illustrated in Figure 8, we inspected the top three users, and saw contrasting behaviors. User B is the official account of a popular local newspaper, while user C belongs to another heating company. We further dug into what the latter said, and found out he was promoting his own heating facilities and encouraged the universities to replace the defect ones in use. However, students reposted the newspaper but not the heating company, because they did not care about this "advertisement". Other discoveries included event starters, who have numerous followers and posted several weibos, and users who were rarely reposted but also actively participated in.

Finally, the following and reposting relations are shown in the entity graph and repost pathway, which confirmed our discoveries.

Besides the findings described above, this tool allows us to dig out various roles of key users. We gained deeper and deeper understanding of this event along the process of exploring.

8 Conclusion And Future Work

In this paper, we proposed a system for interactively identifying key players as well as in-depth exploring their roles in the process of information propagation. We have explored potential features of key players, and provided their formal definition for filtering operations. With the help of multiple linked views, users are able to identify, understand key players and comprehend related events. Cases show our system is effective for key players identification and exploration.

In the future, we will support more entities and attributes to enhance compatibility. Meanwhile, we will further revise the existing framework, because as mentioned in Section 4, many collected features are abstract and still difficult to formalize.

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

Acknowledgments

The authors wish to thank ChinaVis reviewers. This work is supported by NSFC No.61170204 and partially funded by NSFC Key Project No. 61232012 and the National Program on Key Basic Research Project (973 Program) No. 2015CB352500. This work is also supported by PKU-Qihu Joint Data Visual Analytics Research Center.

References

[1] J. Alsakran, Y. Chen, D. Luo, Y. Zhao, J. Yang, W. Dou, and S. Liu. Real-time visualization of streaming text with a force-based dynamic system. *IEEE Computer Graphics and Applications*, 32(1):34–45, 2012.

[2] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2649–2658, Dec 2012.

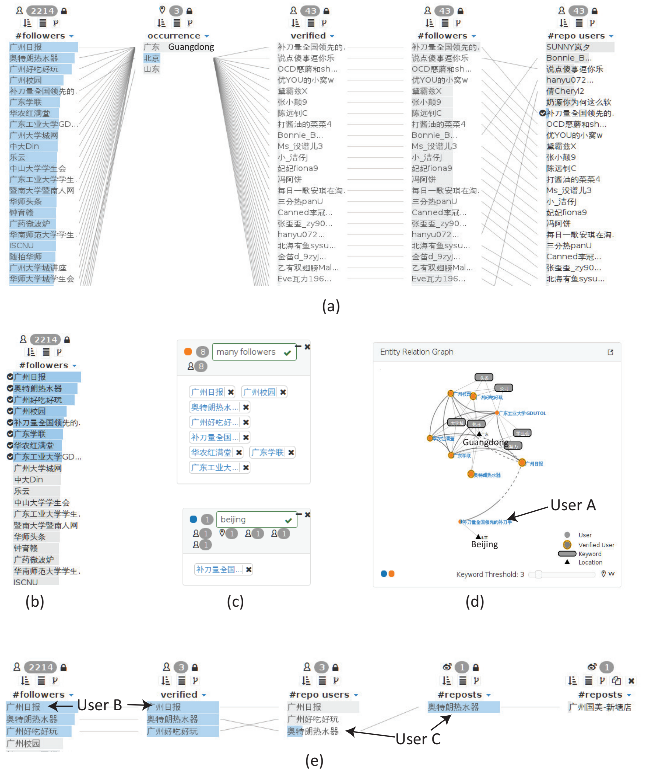

Fig. 8. Screenshots of the hot water supply event illustrating how to select entities step by step and use multiple filters.

[3] 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.

[4] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2412–2421, 2011.

[5] W. Cui, Y. Wu, S. Liu, F. Wei, M. X. Zhou, and H. Qu. Context preserving dynamic word cloud visualization. In *Proceedings of IEEE Pacific Visualization Symposium*, pages 121 –128, 2010.

[6] 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.

[7] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou. Leadline: Interactive visual analysis of text data through event identication and exploration. In *Proc. IEEE Conference on Visual Analytics Science and Technology 2012*, pages 93 – 102, 2012.

[8] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, pages 5–, Washington, DC, USA, 2005. IEEE Computer Society.

[9] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, INFOVIS '05, pages 5–, Washington, DC, USA, 2005. IEEE Computer Society.

[10] N. Henry, J. Fekete, and M. J. McGuffin. Nodetrix: a hybrid visualization of social networks. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1302–1309, 2007.

[11] C.-T. Ho, C.-T. Li, and S.-D. Lin. Modeling and visualizing information propagation in a micro-blogging platform. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 328–335. IEEE, 2011.

[12] M. Hu, S. Liu, F. Wei, Y. Wu, J. Stasko, and K.-L. Ma. Breaking news on twitter. In *CHI '12*, pages 2751–2754, 2012.

[13] S. Kumar, G. Barbier, M. A. Abbasi, and H. Liu. TweetTracker: An analysis tool for humanitarian and disaster relief. In *ICWSM '11*, 2011.

[14] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW '10*, pages 591–600, 2010.

[15] 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.

[16] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, and D. Boyd. The arab spring — the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International Journal of Communication*, 5:1375–1405, 2011.

[17] 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 *IEEE VAST '11*, pages 181 –190, 2011.

[18] 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 *CHI '11*, pages 227–236, 2011.

[19] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 318–322. ACM, 1994.

[20] 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 *WWW '11*, pages 249–252, 2011.

[21] C. Remy, N. Pervin, F. Toriumi, and H. Takeda. Information diffusion on twitter: everyone has its chance, but all chances are not equal. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2013 International Conference on*, pages 483–490. IEEE, 2013.

[22] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704. ACM, 2011.

[23] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW '10*, pages 851–860, 2010.

[24] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 241–250. ACM, 2010.

[25] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. J. Zhu, and R. Liang. Evoriver: Visual analysis of topic coopetition on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1753–1762, 2014.

[26] D. Thom, H. Bosch, S. Koch, M. Wrner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *IEEE PacificVis '12*, pages 41 – 48, 2012.

[27] Y. Tyshchuk, H. Li, H. Ji, and W. A. Wallace. Evolution of communities on twitter and the role of their leaders during emergencies. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 727–733. ACM, 2013.

[28] F. Viégas, M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant, and C. Wren. Google+ripples: A native visualization of information flow. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 1389–1398, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[29] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1763–1772, 2014.

[30] 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.

[31] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. *ICWSM*, 10:355–358, 2010.

[32] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. M. Collins. #fluxflow: Visual analysis of anomalous information spreading on social media. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1773–1782, 2014.