GalaxyExplorer: Influence-Driven Visual Exploration of Context-Specific Social Media Interactions

Xiaotong Liu* The Ohio State University
Srinivasan Parthasarathy† The Ohio State University
Han-Wei Shen‡ The Ohio State University
Yifan Hu§ Yahoo Labs

ABSTRACT

The ever-increasing size and complexity of social networks place a fundamental challenge to visual exploration and analysis tasks. In this study, we present GalaxyExplorer, an influence-driven visual analysis system for exploring users of various theme-aware influence in a social network. GalaxyExplorer reduces the size and complexity of a social network by dynamically retrieving theme-based graphs, and analyzing users' influence and passivity regarding specific context and time in response to disaster events. In Galaxy-Explorer, a galaxy-based visual metaphor is employed to simplify the visual complexity of a large network with a focus+context view. With a client-server implementation, GalaxyExplorer achieves on-the-fly computation, and supports influence analysis and visual exploration with details on demand. We show the usefulness of GalaxyExplorer through a case study in theme-aware influence analysis and visual exploration of influential users' response to Hurricane Sandy on Twitter.

1 INTRODUCTION

Social media have recently become the easiest, fastest and most explosive way to transmit and receive information. The growing ubiquity, communication bandwidth and speed, and cross-platform accessibility of social media have offered both opportunities and challenges for large-scale information sharing and diffusion during disaster management and mass disruption event coordination [23, 17, 20]. A social network can be seen as a large-scale graph, where each node represents a person, and each edge corresponds to a social relationship between the connected people. For instance, an influential Twitter user can connect to thousands of individuals when an edge in the graph reflects a following or retweeting (reposting) connection. Without systematically handling the sheer data size and visual complexity, analysis of such a large graph is very challenging.

While considerable research efforts have been made to visualize and monitor discussion in social media [6, 2, 7, 15, 27], relatively few works have targeted at visual exploration of influential individuals and how they interact with others in a social network. *Influence*, which occurs when a person adapts his/her behavior, attitudes or beliefs based on those of others, has been an important force that directs the dynamics of social media interactions [14]. The further one's messages are propagated in the network by his/her connected users, the more influence he/she may have on others. Equally important is *passivity*, which reflects the barrier to the propagation of messages that is often hard to overcome [18].

In this work, we seek to analyze users' influence and passivity conditioned on specific themes. This context-specific approach has two advantages. First, the massive number of users in a network can be partitioned into smaller theme-based subsets, thus enabling exploring, identifying and analyzing users of interest at a reduced scale that would not be feasible for the entire network. Second, theme-aware analysis can better capture the relative influence and passivity each user has regarding a specific context, after filtering out unrelated noisy discussion.

Based on the theme-aware influence and passivity, we present GalaxyExplorer, an influence-driven visual analysis system for exploring context-specific social media interactions. Our proposed visual-analytic workflow starts by launching a theme query on the massive network data via a client-side search interface. After the server-side search engine retrieves a theme-based graph and analyzes the influence and passivity of every user, the visual interface of GalaxyExplorer (Figure 1) shows a theme-based graph visualization on the client side. To support interactive rendering of the theme-based graphs, we take up a *galaxy* metaphor, where *massive stars* highlight users in focus as large bright points with labels, *dwarf stars* represent users in context as small dim points, and the static or dynamic *light rays* among the stars correspond to the social relationship or message propagation among the users. The analyst can zoom into users of interest and filter out others by interactively brushing the influence and passivity bar charts to specify a cross-filtered range query, within which the selected users are linked and highlighted in the graph immediately. Moreover, the analyst can analyze further with details on demand, by launching a user query on the theme-based graph to retrieve the user's messages covering the theme of interest, or a message query to extract all messages that mentioned specific keywords. In the meanwhile, the analyst can view the social relationship of a specific user, or the propagation of a message during a period of time to understand the discussion diffusion.

We have developed GalaxyExplorer as a web application to release the burden of downloading and maintaining the entire massive social network data by each individual analyst. We show the usefulness and effectiveness of GalaxyExplorer through a case study in theme-aware influence analysis and visual exploration of influential users' response to Hurricane Sandy on Twitter.

The key contributions of this paper are threefold:

- We present a simple yet effective extension for conditioning the computation of influence and passivity of users in a social media network by theme — Theme-aware Influence and Passivity (TIP).
- We present a novel visual-analytic framework for exploring social media interactions driven by the above TIP algorithm — GalaxyExplorer.
- We demonstrate the efficiency and efficacy of GalaxyExplorer and TIP on a real world case study as it relates to emergency response.

2 RELATED WORK

Social media have recently been used for sharing and analyzing information during and after disaster events. Starbird and Palen [21] found that retweets with topical keywords were more likely to be

*e-mail: liu.1952@osu.edu
†e-mail:srini@cse.ohio-state.edu
‡e-mail:hwshen@cse.ohio-state.edu
§e-mail:yifanhu@research.att.com

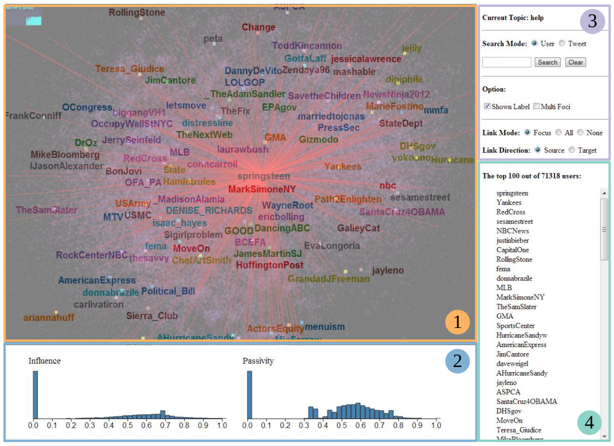

Figure 1: Visual interface of GalaxyExplorer: (1) the main view, (2) the query panel, (3) the control panel, and (4) the list view.

on-topic related to the disaster than nonretweets. Vieweg et al. [23] and Qu et al. [17] found that retweeted information is more likely to contribute to situational awareness. Other studies of information and influence propagation stressed the value of influential individuals as key elements in the propagation of information through the network [5, 1, 8, 25, 18]. Agarwal et al. [1] discovered that the most influential bloggers were not necessarily the most active. Romero et al. [18] proposed a user influence model that also considers the passivity of users when evaluating their influence. To detail with the noisy discussion in online social networks, Ruan et al. [19] combined context and link information in graph structures to discover communities in large networks. Our work aims at analyzing users' influence and passivity while considering the themes as specific contexts as well as the dynamics of the social media interactions in response to disaster events.

Research on social media monitoring and visual analysis have become increasingly active in the fields of information visualization and visual analytics. Dork et al. [6] introduced visual backchannel as a way of following and exploring online conversations about large-scale events. Cuvelier and Aufaure [3] proposed topographic networks of tags, representing a tag cloud with a topographic metaphor to highlight the most important concepts found for a given search on Twitter. Hansen et al. [9] presented a process model to analyze and visualize social media data. Based on offline processing, Twitinfo [16] employed a timeline-based display to highlight peaks of high tweet activity discovered by a streaming algorithm. Whisper [2] traced and visualized the process of information diffusion in social media using a sunflower metaphor. Gansner et al. [7] summarized and visualized tweet clusters using a geographical map metaphor, which was later refined by Liu et al. [15] for better preserving the user's mental map of streaming tweet clusters. Kraft et al. [13] extracted structured representations of Twitter events and visualized key event indicators from Twitter stream. Google+ Ripples [22] showed social media interactions on Google+ with a mix of node-and-link and circular treemap metaphors. Xu et al. [27] analyzed topic computation on social media by integrating ThemeRiver with storyline style visualization. In contrast with the above tools, we target at visual analysis of individuals and show how they influence others' response during critical events, following the visual analytics mantra by Keim et al. [12]. Our system incorporates theme-based influence analysis, large graph drawing and rendering [10], cross-filtered views [24] and detail-on-demand visual exploration.

3 Theme-aware Influence and Passivity

Unlike traditional media, for information to propagate in social media, users need to forward it to others, thus having to actively engage rather than passively read it and seldom act on it. While some users may not have much influence upon the overall social network, they can still play an influential role in the discussion on a specific theme. On the other hand, the size and complexity of a social network can be highly reduced when focusing on a specific context. Therefore, taking into account the topical properties of the network can enhance both quality and scalability of influence analysis.

To achieve this, we adapt a generic influence and passivity (IP) model [18] to capture the theme-based influence each user has in a social network. In the IP model, a user's influence depends on the number of people he/she influences as well as their passivity, and how dedicated the people he/she influences are, while a user's passivity depends on the influence of those who he/she is exposed to but not influenced by, and how much he/she rejects other users' influence compared to everyone else. We seek to analyze the influence and passivity conditioned on specific themes.

Given a social network $G=(V,E)$ with users V and edges E, we first extract a theme-based subgraph $G_t=(V_t,E_t)$ ($V_t \subset V, E_t \subset E$), by retaining the users and edges that are related to a given theme t and filtering out the rest. The theme-based weight w_{ij}^t on the edge e_{ij} represents the ratio of influence that i exerts on j to the total influence that i attempts to exert on j regarding theme t. Taking Twitter network as an example: an edge (i,j) exists if user j reposted a message by user i about theme t, with $w_{ij}^t = n_{ij}^t/s_i^t$, where n_{ij}^t is the number of messages posted by user i that mentioned theme t and reposted by user j, and s_i^t is the total number of messages posted by user i regarding theme t. For every edge $e_{ij} \in E_t$, the *acceptance rate* a_{ij}^t on theme t is defined as:

$$a_{ij}^t = \frac{w_{ij}^t}{\sum_{k:(k,j)\in E_t} w_{kj}^t}. \tag{1}$$

This metric reflects the dedication user j has to user i on theme t. It measures the amount of influence that user j accepted from user i on theme t scaled by the total influence accepted by j from all users. On the other hand, the *rejection rate* r_{ij}^t on theme t is defined as:

$$r_{ij}^t = \frac{1-w_{ij}^t}{\sum_{k:(i,k)\in E_t} (1-w_{ik}^t)}. \tag{2}$$

This metric measures the amount of influence that user j rejected from user i on theme t scaled by the total influence rejected from i by all users.

With these two measures, we employ an iterative process to compute the theme-based influence and passivity for each user, called the *Theme-aware Influence and Passivity (TIP)* algorithm, as shown in Algorithm 1. It first extracts a subgraph regarding a given theme, precomputes theme-based weights and acceptance/rejection rates, and then iteratively computes the theme-based influence and passivity values. Extracting the theme-based graphs (step 1) takes $O(|V|+|E|)$, by simply scanning through all users and edges. Computing the edge weights (step 2), rates and rejection rates (step 3) can be done in $O(|V_t|+|E_t|)$, since the scale is reduced from the entire graph to the theme-based graph. Step 4 – 16 compute the theme-based influence and passivity values, which are mutually updated since they depend on each other — one is considered more influential, if more of his or her messages on theme t have been accepted by the relatively more passive users; one is considered more passive, if more of his or her messages on theme t have been rejected by the relatively more influential users. This computation takes $O(c(|V_t|+|E_t|))$, where c is the total number of iterations for updating these two values, which is a small number according to

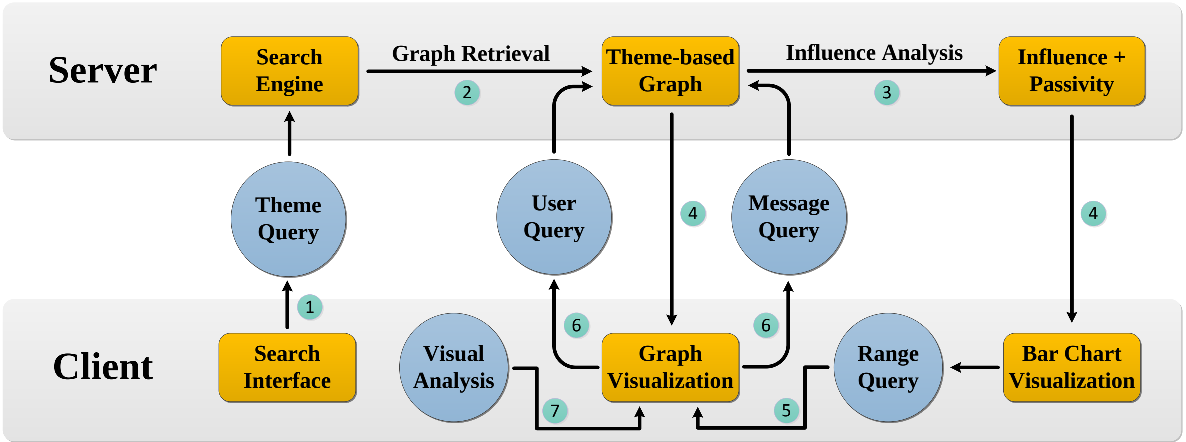

Figure 2: The proposed visual-analytic workflow of GalaxyExplorer.

Algorithm 1 Theme-aware Influence and Passivity (TIP)

Input: An initial network $G = (V, E)$, and a user-specified theme t
Output: The theme-based influence I_i^t and passivity P_i^t for every user $i \in V_t$

1: Extract users $V_t \subset V$ and edges $E_t \subset E$ on theme t to form a theme-based graph $G_t = (V_t, E_t)$
2: Compute the theme-based weight w_{ij}^t for each edge $(i, j) \in E_t$
3: Compute the acceptance rate a_{ij}^t and rejection rate r_{ij}^t for each user $i \in V_t$ according to equations (1) and (2) respectively
4: Initialize influence $I_i^t = 1$ and passivity $P_i^t = 1$ for each user $i \in V_t$
5: **repeat**
6:     **for** each $i \in V_t$ **do**
7:         $P_i^t \leftarrow \sum_{j:(j,i) \in E_t} r_{ji} I_j$
8:     **end for**
9:     **for** each $i \in V_t$ **do**
10:         $I_i^t \leftarrow \sum_{j:(i,j) \in E_t} a_{ij} P_j$
11:     **end for**
12:     **for** each $i \in V_t$ **do**
13:         $P_i^t \leftarrow \frac{P_i^t}{\sum_{k \in V_t} P_k^t}$
14:         $I_i^t \leftarrow \frac{I_i^t}{\sum_{k \in V_t} I_k^t}$
15:     **end for**
16: **until** I_i^t and P_i^t converge

---

the original IP algorithm [18] — the absolute changes of the computed influence and passivity values drop sharply in the first several iterations, and the whole process converges very quickly. Overall, the TIP algorithm takes $O(|V| + |E| + c(|V_t| + |E_t|))$ time.

Since social media are a rapidly changing communication platform, the influence/passivity of a user can be different when taking into account the tweet timestamp. The TIP algorithm can be easily adapted to cope with such dynamics in a theme-based graph, by introducing a time window δ as another constraint when extracting the users and edges on theme t (step 1). In this way, one can compute and compare temporal influence/passivity values in an evolving theme-based graph.

Based on the TIP model, we introduce a metric to evaluate the importance of messages. One tweet is considered more important if it has been reposted by more influential users, and has reached a larger audience. Therefore, with the influences obtained from the TIP algorithm, the *importance* of message k on theme t is defined as:

$$\omega_k^t = \sum_{i \in R_k} (\alpha I_i^t + (1 - \alpha)\lambda), \tag{3}$$

where R_k is the set of users that reposted the message k, λ is a uniform weight for one repost, and $0 \leq \alpha \leq 1$ is a factor to balance the user's specific influence and the uniform weight. A larger α takes into account more of a user's influence when evaluating the importance of a message. On the contrary, when setting $\alpha = 0$ and $\lambda = 1$, the importance is equivalent to the number of reposts for a message.

4 Overview of the Visual-Analytic Workflow

Guided by the classic visual analytics mantra "Analyze first, show the important, zoom, filter and analyze further, details on demand" [12], we propose a carefully designed visual-analytic workflow for exploring social medial interactions driven by theme-aware influence and passivity, as illustrated in Figure 2. Our approach starts from the client side, where an analyst (1) specifies a theme of interest through the search interface. Upon receiving the request, the search engine on the server side (2) retrieves a corresponding graph, (3) analyzes the theme-based influence and passivity, and finally returns the results to the client. The client will be notified when the results are ready, and (4) create a customized visual interface of GalaxyExplorer regarding the specified theme. After that, the analyst can (5) zoom into users of interest while filtering out others by interactively brushing the influence and passivity bar charts to specify a cross-filtered range query, within which the selected users in focus are linked and highlighted in the graph immediately. Furthermore, the analyst can (6) perform analysis with details on demand, by launching a user query on the theme-based graph to retrieve the user's messages covering the theme of interest, or a message query to extract all messages that mentioned specific keywords. At the same time, the analyst can (7) analyze the social relationship of a specific user, or the discussion diffusion of an original message, visualized as an animated propagation in the graph of GalaxyExplorer.

In the next section, we detail the design guidelines identified in these analysis steps, and explain how we achieve them with our interactive, explorative framework GalaxyExplorer.

5 The GalaxyExplorer Framework

When the data are complex and too large to be visualized in a straightforward manner, data analysis should be combined with interactive visualization to enable comprehensive data exploration at both overview and detail levels. Specifically, we outline the following guidelines in designing a scalable visual-analytic framework for exploring social media interactions:

Enable efficient large graph rendering with effective visual metaphor. Even only handling a subset of the entire network, in most cases, the complexity of a theme-based graph is still to much to be displayed. Information overload not only causes visual clutter, but also degrades the system performance. Therefore, an effective visual metaphor that can cope with large graph rendering is highly desirable.

Achieve on-the-fly computation of query search, graph retrieval, influence analysis, and visualization. An efficient visual analysis system requires both powerful computational analysis and visualization. Fast algorithms and techniques are needed to achieve efficient computation for data retrieval, analysis and visualization.

Provide flexible interactions to query users of interest, and view details on demand. In exploration scenarios, the analyst often has no or only vague hypotheses about the users in a graph, and thus needs to interact with the system to navigate through different parts of the graph. A desirable system should allow the analyst to query users of interest, and view details such as reposting relationship and message propagation as they explore the graph.

5.1 Visual Interface

To facilitate an intuitive exploration of the users in a social network, we design GalaxyExplorer based on the metaphor of a galaxy, which has been used in information visualization for its effectiveness in handling visual complexity [26]. In a typical galaxy (Figure 3 (left)), the massive stars are glowing with powerful light. In analogy, in the graph visualization (Figure 3 (right)), the users in focus are highlighted by bright points with name labels. Similarly, the users in context correspond to the dwarf stars in a galaxy, which are relatively dim and small. During exploration, the analyst can interact with the system to examine links among users (analogous to the light rays among the stars) — the static links represent the social relationship among users, while the animated links indicate the propagation process of a specific message. This design reduces the visual complexity of a large graph with a focus+context view, and is suitable for graphs with a large number of users.

Figure 1 shows the visual interface of GalaxyExplorer, which consists of four components: (1) the main view, (2) the query panel, (3) the control panel, and (4) the list view. The main panel visualizes the theme-based graph in a galaxy metaphor. The query panel displays the influence and passivity charts for range query at the overview level, and will transition to a data table when examining and selecting messages at the detail level. The list view shows the top influential users within the specified range query. Several options for user interaction can be adjusted in the control panel. A demonstration of GalaxyExplorer is presented at `http://vimeo.com/69211037`.

5.2 Large Graph Drawing and Rendering

To visualize the graph in this galaxy metaphor, two issues need to be resolved: one is to cope with the large scale of the graph when determining the graph layout, the other is to accomplish a sufficient frame rate when rendering and interacting with the graph.

For graph drawing, force-directed approaches have been widely used to find an efficient placement of the nodes with high quality by minimizing the energy of the physical system. However, even with a multilevel spring-electrical model [10], it still takes minutes or even hours to layout a graph with a large number of nodes and edges. Even if it can handle some smaller theme-based graphs, the positions of the same user across different theme-based graphs can be quite different due to the force-directed model, which makes it difficult for the viewer to locate users of interest across multiple themes. The solve these problems, GalaxyExplorer applies the *scalable force directed placement (sfdp)* method [10] to compute and store the layout of the entire graph on the server side. Whenever the server receives a theme query, it will transmit the layout configuration of the retrieved subgraph to the client. In this way, every user's location is consistent in any theme-based graph, and we do not need to run the layout computation for a second time. To further reduce the data in transmission and thus save the response time on the client side, the floating-point positions of the graph layout generated by sfdp are rounded to integers in $[0, 1000]$ for approximation.

For graph rendering, conventional web graphics such as *Scalable Vector Graphics (SVG)* are not suitable for rendering a graph containing more than thousands of nodes while preserving the reasonable interactivity. Therefore, GalaxyExplorer resorts to *Web Graphics Library (WebGL)*, a JavaScript binding of the *OpenGL ES* that enables low-level imperative graphics rendering based on programmable shaders. When visualizing a galaxy-based graph with around a million nodes, our WebGL implementation achieves more than 30 frames per second (fps), which is sufficient for interactive rendering. In addition, to help the analyst directly view the users in focus during exploration, their names are placed close to the highlighted points as labels in the same color. To reduce visual clutter caused by label overlaps, a repulsive force is employed to remove label overlaps locally while preserving spatial proximity.

5.3 Fast Range Query with Cross Filtering

To allow the analyst to flexibly switch the focus and context for exploring users in a graph, GalaxyExplorer adopts a cross filtering technique to support multi-dimensional range queries for selecting focus visually. Popularized by Chris Weaver [24], cross-filtered views are multiple coordinated views with the following features: (1) each supports selection over the set of unique attribute values in a data column; (2) each data column is paired with a dimensionally appropriate type of view that supports indication of attribute values by selection or navigation; and (3) one can rapidly toggle brushing filters between pairs of views to pose complex drill-down set queries across multiple data columns. In GalaxyExplorer, the data columns correspond to the influence and passivity, and the paired views are bar charts visualizing the distribution of influence/passivity. With cross filtering, one is able to interactively drill down into inter-dimensional relationships buried in the influence and passivity values, and thus flexibly explore the users of various combination of influence and passivity. In the meanwhile, the users that fall within the range query are highlighted immediately in the graph when brushing the two charts. As a result, the cross-filtered view and the graph visualization are dynamically linked for flexible graph exploration.

To achieve fast brushing for cross-filtering thousands and even millions of data records, GalaxyExplorer integrates the graph and

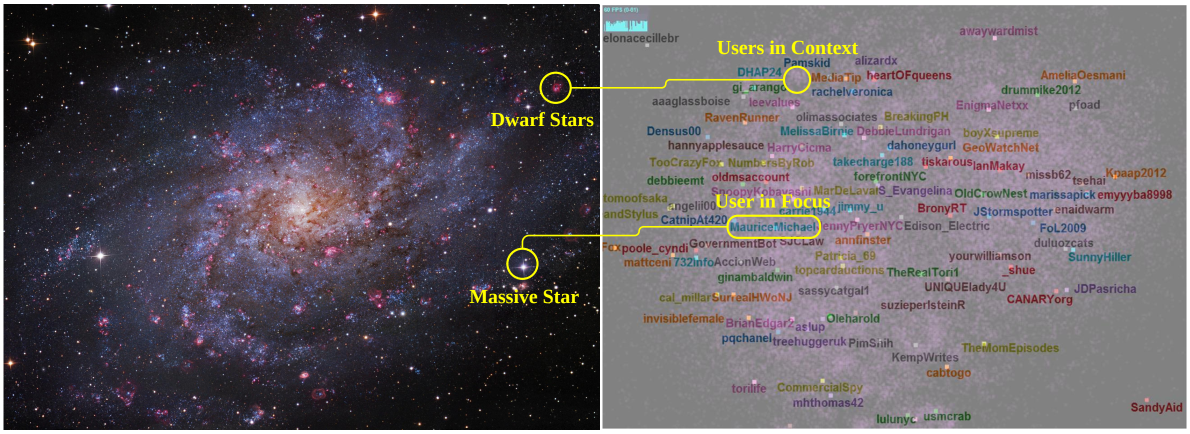

Figure 3: Visualization design of a galaxy-based graph: (left) the M33 Triangulum Galaxy (image by Robert Gendler, Subaru Telescope); (right) the galaxy metaphor in GalaxyExplorer. The massive stars highlight users in focus as large bright points with labels, while dwarf stars correspond to users in context as small dim points.

bar chart visualization with *Crossfilter* [4], a JavaScript library for exploring large multivariate datasets in the browser. We tested cross filtering with more than one million data records, and found it achieved decent interactivity.

5.4 Detail-on-Demand Visual Exploration

To support further comprehensive analysis of users and messages with details on demand, GalaxyExplorer incorporates the following interactive functionalities:

User exploration. The analyst can focus on a user by selecting a label in the graph to view how his/her messages have been reposted by others (denoted as *source links*), or how he/she has reposted others' messages (denoted as *target links*). Multiple foci are supported to allow a direct comparison of different users. In addition, the analyst can search specific users by giving a keyword query to retrieve relevant ones from the server. Upon receiving the results, the query panel of GalaxyExplorer will display a table containing all the matched users and their messages, with which the analyst can perform message exploration.

Message exploration. The analyst can specify a message query to retrieve the messages mentioning keywords of interest, and the query panel of GalaxyExplorer will display the matched messages as well as their corresponding users. The results can be sorted by the names of the users, the dates when the messages were posted, or the importances of the messages. To achieve this, the analyst can simply click on a header of the table, and the retrieved records will be displayed in either decreasing or increasing order of the selected header (a double-click will reverse the order). Furthermore, the analyst can select a specific message (either from the resulting table of the user query or the message query), and an animated transition will illustrate how the message propagated in the graph, where only the users involved in the message propagation are highlighted. The color of the animated trace is mapped to the speed of the message propagation (blue for fast diffusion, red for slow diffusion, and the colors in between for moderate diffusion). At the same time, a time series chart will pop up to illustrate the number of users reached by the selected message during this propagation over time.

6 A Case Study on Hurricane Sandy

To demonstrate the usefulness and effectiveness of GalaxyExplorer, we performed a case study on Twitter users' influence and response to Hurricane Sandy, which was the deadliest and most destructive hurricane of the 2012 Atlantic hurricane season, as well as the second-costliest hurricane in U.S. history. Basically we investigated the following aspects: (1) measuring the computation time to obtain the influence and passivity values for various theme-based graphs, (2) comparing the influence and passivity values across different themes, (3) drilling down to different temporal phases of Hurricane Sandy, and (4) exploring a specific theme in depth.

6.1 Dataset

The dataset used in our studies consists of 3513994 tweets from 2022511 users, which were collected from October 27 to November 7, 2012 — when Hurricane Sandy affected 24 U.S. states. During this period of time, 1167159 users reposted at least one message by others, and 319416 users' messages were reposted. Two users are consider connected if one reposted a message by the other. After filtering out the disconnected users and extracting the largest connected component, we obtained a connected network with 952153 users and 1228880 edges.

6.2 Measuring Computation Time

In the first study, we ran the IP algorithm [18] on the entire graph to obtain the baseline of computation time. Then for every specified theme, we ran the proposed TIP algorithm and recorded the time for extracting the theme-based graph (step 1), precomputing edge weights, acceptance rates and rejection rates (step 2-3), and computing influence and passivity (IP) values (step 4-16), respectively. We selected 20 different keywords that were more or less related to Hurricane Sandy as the themes for this study. Both IP and TIP algorithms were implemented using JavaScript with Node.js [11] for server-side execution. The experiment was conducted on an i7-2600, 3.4 GHz CPU computer.

Figure 4 shows the corresponding computation time. The graphs are ranked by the number of users in decreasing order. The first column (with no theme) stands for the entire graph, which took about 36 seconds to compute the influence and passivity for all users. We found that for the graphs of more than 100000 users (i.e., the entire graph and the theme-based graphs on "Hurricane", "Sandy" and "Storm"), the total computation time is roughly proportional to the graph size, and IP computation dominates the total computation. As the theme-based graphs get smaller, the time for extracting the graph, which does not vary much, contributes more in the total computation, thus resulting in a relatively stable cost. This is consistent with the time complexity of the TIP algorithm. In most cases, it takes a few seconds to run the TIP algorithm with JavaScript implementation. We believe it would take even less time with implementation by low-level programming languages such as C.

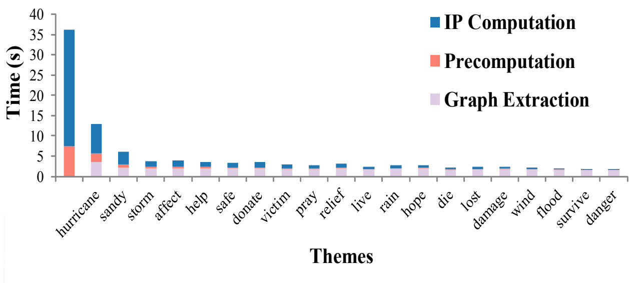

Figure 4: CPU time (in seconds) of graph extraction, precomputation and IP computation for theme-based graphs.

6.3 Comparing Influence and Passivity across Themes

In the second study, we compared the influence and passivity values regarding various themes. We visualized the distribution of users' influence and passivity on different themes using multiple charts to allow side-by-side comparison. When clicking on a point in one chart, the corresponding points reflecting the values of the same user are highlighted in all charts where he/she appeared.

Figure 5 illustrates the aggregated influence and passivity values on the entire graph as well as the theme-based graphs on "Help" and "Danger". The influence values are sorted in decreasing order, while the passivity values in increasing order. When we focused on specific users, we found that the influence and passivity values are often quite diverse. For example, in Figure 5 the influence and passivity of one user is highlighted in a red circle in all charts. As we can see, the selected user in Figure 5 (left) had medium influence in the entire graph, little influence in the graph on "Help", and high influence in the graph on "Danger"; the selected user in Figure 5 (right) was very passive in the entire graph and the graph on "Danger", but very active in the graph on "Help". A possible reason is that the theme-based graphs have filtered out some of the unrelated discussion (since Twitter data are often quite noisy), and thus the IP computation may better reflect the influence and passivity a user has on a specific theme.

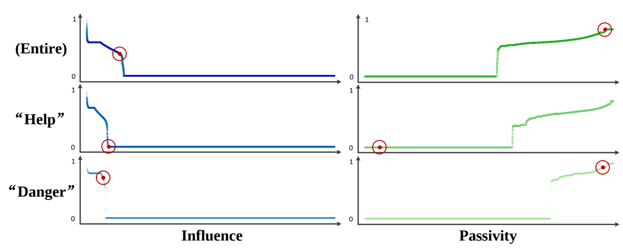

Figure 5: Influence (left) and passivity (right) in the entire graph and theme-based graphs on "Help" and "Danger". The influence and passivity values of a selected user across different graphs are highlighted in a red circle.

6.4 Drilling Down to Temporal Phases of Disaster

In the third study, we partitioned the disaster into different temporal phases — before the disaster (from October 27 to October 29, 2012), during the disaster (from October 29 to November 1, 2012) and after the disaster (from November 1 to November 7, 2012).

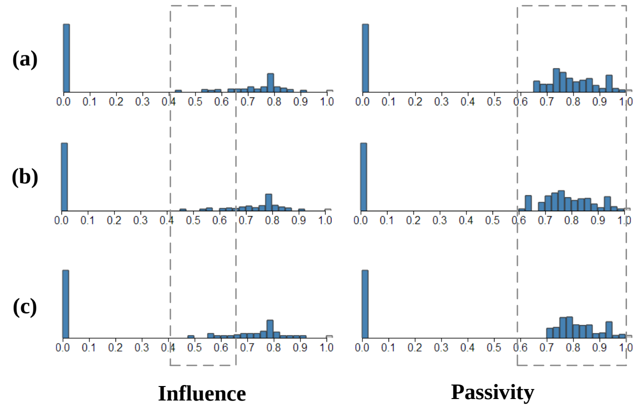

Figure 6: Influence and passivity distributions of the theme "Danger" in different temporal phases: (a) before disaster; (b) during disaster; (c) after disaster.

With GalaxyExplorer, we identified the 20 most influential users at the three phases of the disaster in each theme-based graph. Table 1 shows the ranks of the users in three selected themes "Help", "Donate" and "Relief". In all cases iconic individuals or organizations were the top influencers. For instance, Bruce Sprinsteen (a famous singer-songwriter from New Jersey) was determined to be most influential on the themes of "Help" and "Relief" as his tweets on requests for help and ongoing relief operations reached out to most of the user base and were frequently retweeted. Similarly Bon Jovi (a popular band) led the influence scores for the theme "donate" in large part due to the benefit concert they held to support rescue and relief operations. Donations from concert attendees were highlighted in multiple tweets and retweets. We note that neither of these two individuals were rated in the top 20 when we computed the influence scores on the entire graph (not shown) and moreover these users do not appear in all themes. Similarly, *Yankees* (an American professional baseball team from New York) was ranked among the top 3 influential in both themes "Help" and "Donate", but had much less influence on the theme "Relief". One organization, not surprisingly, that featured in the top influencer lists of all three themes is the American Red Cross (*RedCross*). These findings support our conjecture that users can have quite different relative influence on different themes/topics.

Furthermore, we also found distinct patterns of evolving influences over time regarding a specific theme: (1) the influence of some users (such as *springsteen*, *BonJovi* and *TheDailyEdge*) was quite stable; (2) the influence of some users (such as *HurricaneSandyw*) went up during the disaster but then down afterwards; (3) the influence of some users (such as *RollingStone*) went down during the disaster and then recovered afterwards; (4) some users (such as *StateDept*) were more influential after the disaster; (5) some users (such as *sesamestreet*) became less influential after the disaster, and (6) some users (such as *MoveOn*) became influential only after the disaster. The *RedCross* was relatively more influential with respect to "Help" and "Relief" than "Donate". Additionally, we found that the actual influence score of the Red Cross to rise (not shown) as we moved to the "during" and "aftermath" stages of the disaster (where relief activities started taking center stage).

There were also several surprising entries that are also highlighted by our study. *sesamestreet*, an American children's television series, posted *In case you missed it heres our hurricane toolkit. It can help your children understand what is going on*. This tweet had a great reach and influence within the "Help" topic accounting for its high influence score. *MoveOn*, a progressive organization made a late entry to the top influence list post disaster. After

Table 1: Theme matters — the 20 most influential users at different temporal phases of disaster in three selected themes “Help”, “Donate” and “Relief”. Selected users are highlighted in color at different phases.

HELP			DONATE			RELIEF		
Before	During	After	Before	During	After	Before	During	After
springsteen	springsteen	springsteen	BonJovi	BonJovi	BonJovi	springsteen	springsteen	springsteen
Yankees	Yankees	RedCross	edshow	digiphile	edshow	TheDailyEdge	TheDailyEdge	TheDailyEdge
RedCross	RedCross	Yankees	Yankees	Yankees	Yankees	RedCross	RedCross	RedCross
sesamestreet	sesamestreet	NBCNews	digiphile	edshow	WillFerrell	BarackObama	thinkprogress	thinkprogress
NBCNews	NBCNews	sesamestreet	WillFerrell	HurricaneSandyw	digiphile	thinkprogress	BarackObama	BarackObama
justinbieber	justinbieber	justinbieber	AdamSchefter	WillFerrell	AdamSchefter	WillFerrell	digiphile	WillFerrell
RollingStone	CapitalOne	RollingStone	justinbieber	TheSamSlater	justinbieber	loudobbsnews	loudobbsnews	loudobbsnews
CapitalOne	RollingStone	CapitalOne	HurricaneSandyw	justinbieber	SportsCenter	TeaPartyCat	WillFerrell	SIPeterKing
MarkSimoneNY	donnabrazile	MarkSimoneNY	SportsCenter	SportsCenter	StateDept	donnabrazile	donnabrazile	AdamSchefter
donnabrazile	fema	donnabrazile	RedCross	AdamSchefter	RedCross	AdamSchefter	TeaPartyCat	AmericanExpress
MLB	MLB	MLB	StateDept	StateDept	AmericanExpress	SIPeterKing	AdamSchefter	TeaPartyCat
GMA	MarkSimoneNY	GMA	TeaPartyCat	RexHuppke	TeaPartyCat	AmericanExpress	SIPeterKing	donnabrazile
JimCantore	TheSamSlater	JimCantore	AmericanExpress	RedCross	GMA	digiphile	AmericanExpress	digiphile
fema	GMA	SportsCenter	GMA	AmericanExpress	HurricaneSandyw	kgosztola	TUSK81	RollingStone
SportsCenter	HurricaneSandyw	daveweigel	RexHuppke	TeaPartyCat	CapitalOne	TUSK81	MLB	MonicaCrowley
AmericanExpress	SportsCenter	AmericanExpress	CapitalOne	GMA	MySportsLegion	RollingStone	MonicaCrowley	kgosztola
ASPCA	AmericanExpress	fema	MySportsLegion	CapitalOne	Arianna8927	NewYorkPost	RexHuppke	TUSK81
HurricaneSandyw	JimCantore	jayleno	donnabrazile	MySportsLegion	SIPeterKing	MonicaCrowley	NewYorkPost	NewYorkPost
daveweigel	daveweigel	MoveOn	SIPeterKing	SIPeterKing	donnabrazile	CapitalOne	RollingStone	CapitalOne
jayleno	jayleno	HurricaneSandyw	Arianna8927	toddstarnes	toddstarnes	EvaLongoria	kgosztola	MoveOn

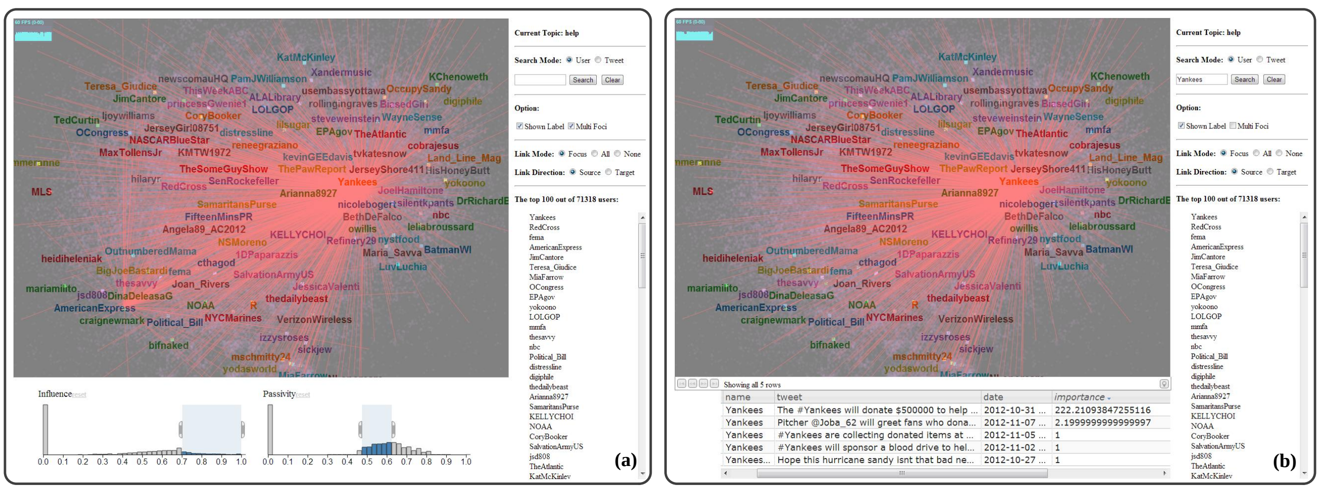

Figure 7: Visual exploration of the graph on theme “Help”: (a) displaying the source links from the user *Yankees* and *AmericanExpress* to others within the influence range Q_i = [0.7, 1.0] and the passivity range Q_p = [0.475, 0.625]; (b) focusing on the tweets by *Yankees* in decreasing order of importance.

drilling down using GalaxyExplorer — the specific tweet that contributed greatly to this score can be attributed to political innuendo especially in light of the proximity of the US presidential elections to this event — is described below: *New Jersey GovChristie on hurricane Sandy response: The President deserves great credit*. Similarly *TheDailyEdge*, a political Twitter channel, tweeted *#Romney: I will bring big changes to America. For example, no more federal hurricane relief*, and had a major influence. Overall, the top influencers on the theme “Relief” seemed to be dominated by political media and innuendo. Contextually this makes sense since the 57th U.S. presidential election was just a week after Hurricane Sandy. All of these highlight the need for theme-aware influence analysis.

In addition, Figure 6 shows the distributions of influence and passivity values with respect to different temporal phases of the theme “Danger” from GalaxyExplorer. We can see that the distributions vary across the time periods (particularly within the range of the dashed lines). Interestingly, the passivity distributions were more diverse than the influence distributions regarding the partitioned temporal phases.

6.5 Exploring a Specific Theme in Depth

In the final study, we explored Twitter users’ influence and passivity in depth regarding the graph on theme “Help”.

To start with, we explored the users with high influence and medium passivity and how they stepped forward regarding Hurricane Sandy. We brushed the influence chart with a range query Q_i = [0.7, 1.0], and the passivity chart with a range query Q_p = [0.475, 0.625]. Figure 7 (a) shows the corresponding view of GalaxyExplorer. From the list panel, we observed that *Yankees* and *AmericanExpress* were the first and fourth top influential users within the cross-filtered ranges. We selected these two users in the graph to see their source links. From the main view of Figure 7(a), we can see that *Yankees* had more source links than the *AmericanExpress*, indicating that the messages posted by *Yankees* received more at-

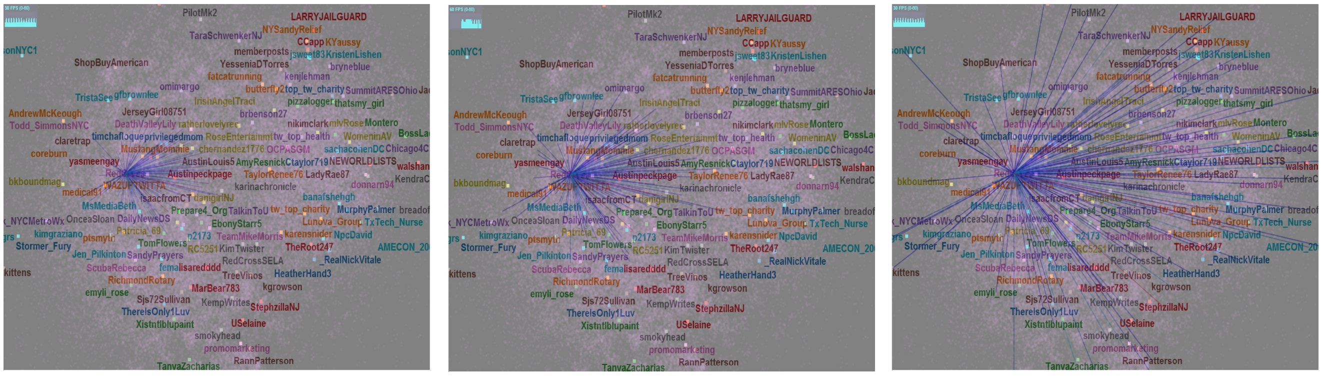

Figure 8: Selected frames for the animated propagation of a specific tweet by *RedCross* in the graph on theme "Help", from November 2 to November 5, 2012. The colors of the animated traces are mapped to the speeds of the message propagation (blue for fast diffusion, red for slow diffusion, and the colors in between for moderate diffusion).

tention. To view more details, we searched the tweets posted by *Yankees* through the control panel, and the results were displayed in the query panel of Figure 7(b). We found one of the tweets by *Yankees* on October 31, 2012 was *The #Yankees will donate $500,000 to help the tri-state area recover from Hurricane Sandy*, whose importance value was 222.21, and was reposted by more than 1000 users in the graph on "Help". Similarly, we found a tweet on November 3, 2012 by *AmericanExpress* saying *Want to help Hurricane #Sandy relief efforts? Cardmembers can donate #MembershipRewards pts to charities*, with the importance value 29.81 and around 160 retweets. It may imply that the organizations (such as the Yankees) in the strongly affected areas (such as New York) are more likely to reach a larger audience through retweets in response to the disaster.

On the other hand, during exploration of the users with low influence (Q_i = [0, 0.2]) and high passivity (Q_p = [0.775, 1.0]), we observed that there were no source links but only target links among the retrieved users within the range queries. This reveals that the passive and least influential users received no retweets (source links) but only forwarded others' tweets (target links) when discussing the theme "Help".

After that, we explored the discussion diffusion of a specific tweet *Interactive map for Hurricane Sandy relief efforts: http://bit.ly/PwbCJ7* by *RedCross*, which was posted on November 2, 2012 and kept being reposted until November 5, 2012. Figure 8 shows three selected frames for the animated propagation of this tweet, with growing propagation traces. We observed that most of the traces are in blue, indicating that most of its retweets occurred very quickly after its post.

7 Conclusion and Future Work

We present GalaxyExplorer, an influence-driven visual analysis system for exploring context-specific social media interactions. We have designed, implemented, and demonstrated an visual-analytic framework for theme-aware influence analysis. A galaxy-based visual metaphor is employed to simplify the visual complexity of a large network with a focus+context view. With a client-server implementation, GalaxyExplorer achieves on-the-fly computation, and supports influence analysis and visual exploration with details on demand. Our case study on Hurricane Sandy has shown that our theme-aware approach reflects the diversity of influence in social media interactions with respect to specific context and time, which would not be revealed by the entire social network.

In the future, a thorough user study and more case studies will be conducted to further investigate the usability of our system. We will also employ shrinkage estimation for better handling the sparsity of theme-based influence and passivity analysis, and study how to select a significant theme for exploration regarding the contents of the discussion as well as the analyst's interest.

References

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *WSDM*, 2008.

[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. *Transactions in Visualization and Computer Graphics*, 2012.

[3] E. Cuvelier and M.-A. Aufaure. A buzz and e-reputation monitoring tool for twitter based on galois lattices. In *International conference on Conceptual structures for discovering knowledge*, 2011.

[4] J. Davies. Crossfilter: a javascript library for exploring large multivariate datasets in the browser., 2012.

[5] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, 2001.

[6] M. Dork, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *Transactions in Visualization and Computer Graphics*, 2010.

[7] E. R. Gansner, Y. Hu, and S. North. Visualizing streaming text data with dynamic graphs and maps. In *Graph Drawing*, 2012.

[8] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *ACM international conference on Web search and data mining*, 2010.

[9] D. L. Hansen, D. Rotman, E. Bonsignore, N. Milic-Frayling, E. M. Rodrigues, M. Smith, and B. Shneiderman. Do you know the way to sna?: A process model for analyzing and visualizing social media network data. In *International Conference on Social Informatics*, 2012.

[10] Y. Hu. Efficient and high quality force-directed graph drawing. *The Mathematica Journal*, 2005.

[11] Joyent. Node.js, http://nodejs.org/.

[12] D. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in visual data analysis. In *Information Visualization*, 2006.

[13] T. Kraft, D. Wang, J. Delawder, W. Dou, Y. Li, and W. Ribarsky. Less after-the-fact: Investigative visual analysis of events from streaming twitter. In *IEEE Symposium on Large-Scale Data Analysis and Visualization*, 2013.

[14] R. Leenders. Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix. *Social Networks*, 2002.

[15] X. Liu, Y. Hu, S. North, and H.-W. Shen. Compactmap: A mental map preserving visual interface for streaming text data. In *IEEE International Conference on Big Data*, 2013.

[16] 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 *ACM CHI Conference on Human Factors in Computing Systems*, 2011.

[17] 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 *ACM conference on Computer supported cooperative work*, 2011.

[18] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *WWW*, 2011.

[19] Y. Ruan, D. Fuhry, and S. Parthasarathy. Efficient community detection in large networks using content and links. In *WWW*, 2013.

[20] A. Sarcevic, L. Palen, J. White, K. Starbird, M. Bagdouri, and K. Anderson. "beacons of hope" in decentralized coordination: learning from on-the-ground medical twitterers during the 2010 haiti earthquake. In *ACM conference on Computer Supported Cooperative Work*, 2012.

[21] K. Starbird and L. Palen. Pass it on?: Retweeting in mass emergencies. In *Information Systems for Crisis Response and Management Conference*, 2010.

[22] 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 *WWW*, 2013.

[23] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *ACM CHI Conference on Human Factors in Computing Systems*, 2010.

[24] C. Weaver. Cross-filtered views for multidimensional visual analysis. *Transactions in Visualization and Computer Graphics*, 2010.

[25] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive infuential twitterers. In *WSDM*, 2010.

[26] J. Wise, J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Information Visualization*, 1995.

[27] 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. *IEEE Visual Analytics Science and Technology*, 2013.