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

Xiaotong Liu
The Ohio State University
Columbus, OH
liu.1952@osu.edu

Srinivasan Parthasarathy
The Ohio State University
Columbus, OH
parthasarathy.2@osu.edu

Han-Wei Shen
The Ohio State University
Columbus, OH
shen.94@osu.edu

Yifan Hu
Yahoo Labs
New York, NY
yifanhu@yahoo.com

ABSTRACT

The ever-increasing size and complexity of social networks place a fundamental challenge to visual exploration and analysis tasks. In this paper, we present *GalaxyExplorer*, an influence-driven visual analysis system for exploring users of various influence and analyzing how they influence others 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 themes and dynamics in response to disaster events. In GalaxyExplorer, a galaxy-based visual metaphor is introduced to simplify the visual complexity of a large graph with a focus+context view. Various interactions are supported for visual exploration. We present experimental results on real-world datasets that show the effectiveness of GalaxyExplorer in theme-aware influence analysis.

1. INTRODUCTION

The growing ubiquity, communication bandwidth, 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 [8, 11]. Research efforts have been made to visualize and monitor discussion in social media [6] and explore how public posts are shared in a social network [10]. *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 [1, 2, 3, 5, 12]. 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 [9].

In this work, we seek to analyze users' influence and passivity conditioned on specific themes. 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 shows a theme-based graph visualization on the client side. A *galaxy* metaphor is introduced to simplify the visual complexity of a large graph. With GalaxyExplorer, the analyst can view users of interest and filter out others by interactively specifying 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 show the effectiveness of GalaxyExplorer in theme-aware influence analysis regarding Hurricane Sandy on Twitter.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742828.

2. THEME-AWARE INFLUENCE ANALYSIS

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.

To capture the theme-based influence each user has in a social network, we adapt a generic influence and passivity (IP) model [9]. 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. 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.

Given a social network $G = (V, E)$ with users V and edges E, we first extract a theme-based subgraph $G_t = (V_t, E_t)$ $(V_t \subset V, E_t \subset E)$,

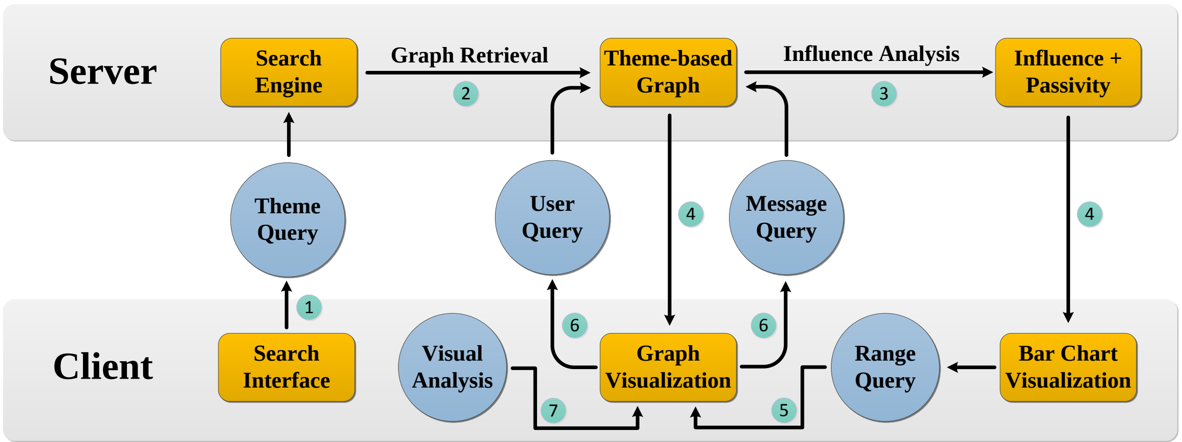

Figure 1: The visual-analytic workflow of GalaxyExplorer.

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.

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 $i \in V_t$

Extract users $V_t \subset V$ and edges $E_t \subset E$ on theme t to form a theme-based graph $G_t = (V_t, E_t)$
Compute the theme-based weight w_{ij}^t for each edge $(i,j) \in E_t$
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
Initialize influence $I_i^t = 1$ and passivity $P_i^t = 1$ for each user $i \in V_t$
repeat
  for each $i \in V_t$ **do**
    $P_i^t \leftarrow \sum_{j:(j,i)\in E_t} r_{ji} I_j$
  end for
  for each $i \in V_t$ **do**
    $I_i^t \leftarrow \sum_{j:(i,j)\in E_t} a_{ij} P_j$
  end for
  for each $i \in V_t$ **do**
    $P_i^t \leftarrow \frac{P_i^t}{\sum_{k\in V_t} P_k^t}$
    $I_i^t \leftarrow \frac{I_i^t}{\sum_{k\in V_t} I_k^t}$
  end for
until I_i^t and P_i^t converge

3. VISUAL-ANALYTIC WORKFLOW

Guided by the classic visual analytics mantra "Analyze first, show the important, zoom, filter and analyze further, details on demand" [4], we propose a carefully designed visual-analytic workflow for exploring social medial interactions driven by theme-aware influence and passivity, as illustrated in Figure 1. 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.

4. VISUALIZATION FRAMEWORK

When the data are complex and too large to be visualized in a straightforward manner, data analysis should be combined with in-

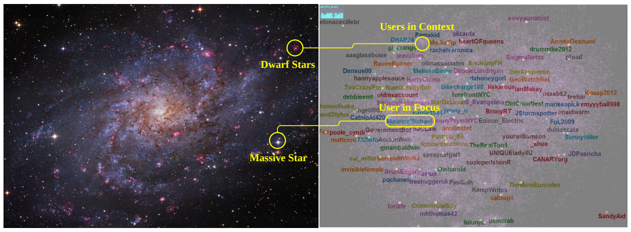

Figure 2: Visualization design of a galaxy-based graph: (left) the M33 Triangulum Galaxy (image by Robert Gendler, Subaru Telescope); (right) the galaxy metaphor in GalaxyExplorer.

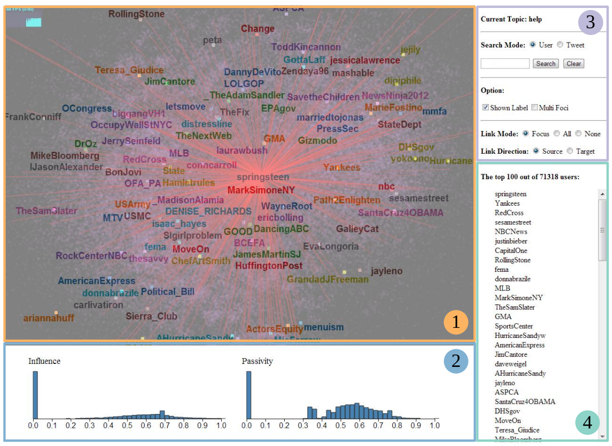

Figure 3: Visual interface of GalaxyExplorer: (1) the main view, (2) the query panel, (3) the control panel, and (4) the list view.

teractive visualization to enable comprehensive data exploration at both overview and detail levels.

4.1 Visualization Design

To facilitate an intuitive exploration of the users in a social network, we design GalaxyExplorer based on the metaphor of a galaxy. In a typical galaxy (Figure 2 (left)), the massive stars are glowing with powerful light. In analogy, in the graph visualization (Figure 2 (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 3 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.

4.2 Interactions

GalaxyExplorer allows the analyst to query users of interest, and view details such as reposting relationship and message propagation as they explore the graph.

Fast range query with cross filtering. 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.

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, or how he/she has reposted others' messages. 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 importance of the messages. Furthermore, the analyst can select a specific message, 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. 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.

5. AUDIENCE EXPERIENCE

We plan a demonstration that leads audience participants through various elements of GalaxyExplorer. A screencast of GalaxyExplorer is available at `http://vimeo.com/69211037`. Our pre-loaded datasets consist of over 3 million tweets from 2 million Twitter users. The datasets were collected from October 27 to November 7, 2012 — when Hurricane Sandy affected 24 U.S. states. The current GalaxyExplorer system is developed as a web application using HTML5 and JavaScript.

With GalaxyExplorer, participants will experience the workflow presented in Figure 1. Participants will be given the opportunity to specify themes of interest through the search interface of GalaxyExplorer. The search engine on the server side will retrieve the corresponding graphs and analyze the theme-based influence and passivity. The participants will then be able to view and compare the Twitter users' influence and passivity values regarding various themes of interest. Through the graphical interface of GalaxyExplorer, participants can explore Twitter users of interest while filtering out others by interactively brushing the influence and passivity bar charts. Participants will also be able to see how tweets have been retweeted in the network. Explicit feedback from audience participants will be sought to rate and evaluate the system along the traditional axes of quality, efficiency and usability.

6. POTENTIAL IMPACT

The ability to explore the heterogeneous aspects of Twitter users' influence and passivity with GalaxyExplorer can help study 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.

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. Neither of these two individuals were rated in the top 20 when we computed the influence scores on the entire graph 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*).

Furthermore, distinct patterns of evolving influences over time can also be found regarding a specific theme: (1) the influence of *springsteen*, *BonJovi* and *TheDailyEdge* was quite stable; (2) the influence of *HurricaneSandyw* went up during the disaster but then down afterwards; (3) the influence of *RollingStone* went down during the disaster and then recovered afterwards; (4) *StateDept* was more influential after the disaster; (5) *sesamestreet* became less influential after the disaster, and (6) *MoveOn* became influential only after the disaster. Further results and analysis details that we lacked space to describe can be found in [7].

In conclusion, our system may allow analysts to compare the users' influence in social media across themes, drill down to temporal phases of critical events in disaster management, and explore detailed messages to gain insights about the diffusion of influence in social media interactions.

7. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their comments. This work was supported in part by NSF grants IIS-1111182, IIS-1111118, IIS-1250752, IIS-1065025, and US Department of Energy DE-SC0007444, DE-DC0012495, program manager Lucy Nowell.

8. REFERENCES

[1] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *International Conference on Web Search and Data Mining*, 2008.

[2] P. Domingos and M. Richardson. Mining the network value of customers. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2001.

[3] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *International Conference on Web Search and Data Mining*, 2010.

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

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

[6] 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.

[7] X. Liu, S. Parthasarathy, H.-W. Shen, and Y. Hu. Galaxyexplorer: Influence-driven visual exploration of context-specific social media interactions. In *OSU Technical Report Series*, 2014.

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

[9] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *World Wide Web*, 2011.

[10] 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 *World Wide Web*, 2013.

[11] 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 Conference on Human Factors in Computing Systems*, 2010.

[12] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive infuential twitterers. In *International Conference on Web Search and Data Mining*, 2010.