Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time

Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu, *Member, IEEE*

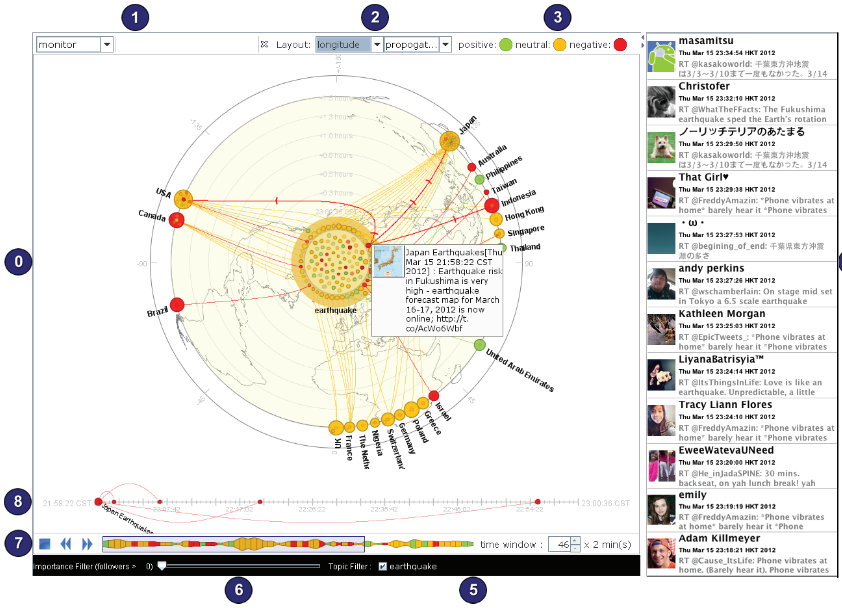

Fig. 1. The figure shows a diffusion of information on Twitter regarding a recent 6.8 magnitude earthquake and a series of aftershocks and tsunamis that hit the northern coast of Hokkaido island (a demo on youtube: http://www.youtube.com/watch?v=aHYU66Z-4FM). The event caught global attention because the location was one of the areas in northeast Japan devastated by last year's disaster. Other countries, including Australia, were initially concerned about the Pacific-wide tsunami threat triggered by the earthquake. Our visualization "Whisper" traced this event in real time. In this figure, original tweets are placed at the center of a circle and pathways are created to connect to geo-groups once the tweets got re-broadcasted (retweeted) by the groups. The retweeting activity is shown as a sequence of color-coded retweet glyphs moving along pathways indicating the timing and sentiments of the retweets. The numbering annotations from 0 to 8 correspond to major design components and functionality which will be described in detail in the paper.

Abstract—When and where is an idea dispersed? Social media, like Twitter, has been increasingly used for exchanging information, opinions and emotions about events that are happening across the world. Here we propose a novel visualization design, "Whisper", for tracing the process of information diffusion in social media in real time. Our design highlights three major characteristics of diffusion processes in social media: the temporal trend, social-spatial extent, and community response of a topic of interest. Such social, spatiotemporal processes are conveyed based on a sunflower metaphor whose seeds are often dispersed far away. In Whisper, we summarize the collective responses of communities on a given topic based on how tweets were retweeted by groups of users, through representing the sentiments extracted from the tweets, and tracing the pathways of retweets on a spatial hierarchical layout. We use an efficient flux line-drawing algorithm to trace multiple pathways so the temporal and spatial patterns can be identified even for a bursty event. A focused diffusion series highlights key roles such as opinion leaders in the diffusion process. We demonstrate how our design facilitates the understanding of when and where a piece of information is dispersed and what are the social responses of the crowd, for large-scale events including political campaigns and natural disasters. Initial feedback from domain experts suggests promising use for today's information consumption and dispersion in the wild.

Index Terms—Information visualization, Information diffusion, Contagion, Social media, Microblogging, Spatiotemporal patterns

1 Introduction

Over the last couple of years, the Internet and social media, most notably Twitter and Facebook, have revolutionized the way news is disseminated and the way we consume it. This includes the series of protests erupted across the Middle East, the news of bin Laden's death, and the reactions to potentially disastrous situations like earthquakes. Before CNN confirmed that U.S. Navy SEALS killed Osama bin Laden, millions had already rapidly dispersed the information through their Twitter and Facebook pages [7]. A New Yorker may have read about the DC earthquake before s/he felt the shake in New York. Thanks to the instant publishing capabilities of social media sites, individuals as well as news outlets are both broadcasting and watching events as they unfold in real time across the globe.

- *Nan Cao and Huamin Qu are with the Hong Kong University of Science and Technology, e-mail: nancao@gmail.com,huamin@cse.ust.hk*
- *Yu-Ru Lin and David Lazer are with Northeastern University and Harvard University, e-mail: yuruliny,davelazer@gmail.com.*
- *Xiaohua Sun is with TongJi University, e-mail: xsun@tongji.edu.cn.*
- *Shixia Liu is with Microsoft Research Asia, e-mail: liushixia@gmail.com*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

1077-2626/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

The tweeting on bin Ladin has been coined the "CNN moment" of Twitter [7]; yet events may attract public attention in numerous ways. On social media sites, some events spread rapidly, while others may be pitched out slowly. Local news or community interests, such as the Wisconsin union protests, are likely to be trending after re-tweeting by celebrities or journalists. How can we trace the different processes of information diffusion in social media?

In this paper, we propose a novel visual design termed "Whisper" to fulfill the need for tracing information diffusion processes in social media, in a real time manner. Whisper, aimed at answering *when, where, and how an idea is dispersed*, is designed to highlight three major characteristics of information diffusion: temporal trend, social-spatial extent, and community response to a topic of interest.

We visualize such social, spatiotemporal processes based on the metaphor of a "sunflower" whose seeds are often dispersed far away. Our design incorporates several novel visual ingredients: a hierarchical social-spatial layout, the pathways of information flow, and a dynamic diffusion series. We use data collected from the social media site Twitter, where information flow may be observed via the "retweet" (RT) feature[1] – a user takes a tweet posted by someone else and rebroadcasts the same message to his or her followers. In Whisper, original tweets are placed inside a circle in the center. A pathway is created to connect a tweet to a user community once it gets retweeted by that community.A sequence of color-coded retweet glyphs moving along the pathways indicates the timing and sentiments of the retweets (see Fig. 1 and Fig. 2). Since a retweet itself does not add new content or sentiment to the original tweet, the sentiment information can be derived just from the original tweet. Thus the retweeting action can be interpreted as an agreement with the sentiment of the original tweet. In this paper we utilize the geographical information to group users into geo-communities; however, different ways of grouping, for example, grouping based on users' profile information or network community detection [35], are also possible. The grouping based on geo-information conveniently enables us to show both temporal and spatial traces of retweets. Groups can be zoomed in to show the next level of grouping. We use an efficient magnetic line-drawing algorithm to create multiple pathways such that the temporal and spatial patterns can be identified even for a bursty event. A dynamic diffusion series is used to highlight the key roles such as opinion leaders in the diffusion process. Our implementation allows users to monitor an ongoing topic of interest in real time.

The key contribution of this work is that, with these visual ingredients, users are now able to characterize different diffusion processes on Twitter, including:

- **Collective responses**: The collective responses to a given topic are summarized through the flow of retweets (represented as "diffusion pathways") and the node sizes of communities, with colors indicating the aggregated sentiments of the communities.
- **Multistep information flow**: Original tweets or retweets that initiate significant amount of retweets can be observed through the density of tweets around the center, the density of diffusion pathways and the focused diffusion series, which help identify the opinion leaders in a multistep information flow [32].
- **Temporal heterogeneity**: The density and orientation of moving retweet glyphs capture whether a given topic is a short burst or a long-lasting trend, and whether the burst or trend differs for different communities.

The rest of this paper is organized as follows. We first discuss related work in section 2, followed by the details of our visualization design in section 3. We then describe the system architecture and its usage scenario in section 4 followed by the system implementation details in section 5. We present case studies and expert interviews to justify the usefulness of our design in section 6. Finally, in section 7 we conclude with a discussion and future directions.

[1]https://support.twitter.com/articles/77606-what-is-retweet-rt

2 Related Work

Existing works that are most relevant to ours fall into three categories: information diffusion analysis, social media monitoring, and visualization for illustrating information diffusion patterns. In this section, we briefly summarize the work in each category.

2.1 Analysis of Information Diffusion

Propagation of information through social networks has been a long-term subject extensively studied by researchers in the field of sociology [16, 33, 50]. With the advent of online social networks, blogs, and micro-blogs, information diffusion has drawn considerable research attention in the area of data mining. Extensive researches have been conducted to analyze information diffusion of the social awareness streams by modeling [22, 27, 29], predicting [52], and measuring [43, 53] the flow and structural patterns of information diffusion processes. In contrast with these works, our research utilizes the visualization techniques to monitor the process of information diffusion and reveals the heterogeneity among the diffusion patterns of topics.

2.2 Social Media Monitoring Tools

Various online tools are designed to monitor social media data [31, 45]. Although few of them have been formally published in the literature, they have been widely used and provide powerful functional extensions and complements to current primary social media like Twitter and Facebook. Different tools are designed to monitor different information. Some tools aim at monitoring topic trends or people, such as Tweetstats [4] and Wefollow [6]. Cuvelier and Aufaure's work [21] monitors users' electronic "reputation" over time. Some more sophisticated tools [23,36] are designed to monitor events that have occurred in micro-blogs by topic, geo-space, and time. Mislove et al. [9] monitor the spatiotemporal sentiment patterns in twitter using a density-preserving cartogram visualization [24]. The tweet sentiments are summarized state-by-state and represented by colors and shapes in the cartogram of United States. However, the visualization is not applicable to trace information diffusion.

Some traditional visualization techniques such as map visualizations, statistical diagrams [44], tag clouds [47], dashboards [26], theme rivers [28], and treemaps [46] have been used in monitoring tools. For example, monitter [2] and twinitor [5] provide dashboard views that allow users to monitor and compare different topics simultaneously. Trendistic [3] represents the keyword frequencies overtime in an interactive frequency chart. Tweetstats and the Twitter StreamGraph [1] monitor and summarize the topic trends as tag clouds and theme river respectively. TwitInfo [36] uses coordinated views that integrate map visualization with statistical graphics for event detection. ThemeCrowds [12,15] represents user groups as cells in treemaps with topics discussed by each group laid over the corresponding cell in the form of a tag cloud. In general, there are also visualization approaches from other domains, which combine hierarchical information [30] with other data characteristics and additional relational information, such as TimeRadarTrees [17] and StarGate [39]. However, these approaches do not consider a highly dynamic data input and do not focus on capturing the collective activities in social media.

In contrast with the above tools, Whisper is designed for tracing the dynamics of information diffusion. Furthermore, Whisper provides an efficient filtering mechanism that automatically detects important posts based on users' retweeting behavior [34, 42].

2.3 Visualizations of Information Diffusion Patterns

More and more visualizations have been proposed to reveal the diffusion patterns in micro-blogs. Many of them directly plot the propagation network over social communities to reveal topological patterns. For example, [34, 51] introduce retweet trees (networks) that show the retweeting behavior of twitter users for diffusion pattern comparisons. However, the spatiotemporal patterns of topic spreading remain unclear in these designs. The representation of information diffusion requires displaying interrelated people, space and time simultaneously. Some visualizations like [29] introduce coordinated views [20] to depict information diffusion from multiple perspectives. However, even though coordinated, the separated views still lead to visual discontinuities and break users' mental maps. Instead of using coordinated views, recent applications [8, 10, 11] have integrated the propagation of various information into one single visualization.

Cascade [8] monitors user behavior and sharing events of new articles on twitter. It provides a 3D visualization which can be smoothly transformed into different views for exploring different aspects of information. Google+ Ripples [10] monitors and depicts how a single post spreads over the user network and its communities. Riot Rumours [11] employs a circle packing based design that reveals how unsubstantiated rumors spread via twitter. It provides a detailed representation of individual tweets and topics and shows their sentiments by color.

All these applications represent rich information through various visual encodings to avoid multiple coordinated views [20]. However, complex encodings may lead to a longer learning curve and hence the amounts of information that can be represented in these applications are limited (e.g. focusing only on sentiments or focusing on diffusion triggered by a single post). The goal of Whisper is to provide a more complete story about information diffusion over time and among people and space. In contrast with the above designs, we design Whisper to leverage the advantages of visual encodings and coordinated views – it traces the dynamics of information diffusion by using an intuitive sunflower visual metaphor to construct a narrative among topic, people, time, and location within one single visualization, with coordinated views such as diffusion series to facilitate detailed understanding about a particular element.

3 VISUALIZATION DESIGN

In this section we discuss the design rationales that influence the design of Whisper (section 3.1). We then present Whisper's visual encoding methodology (section 3.2). Finally, we describe the interaction designs to support an exploration of diffusion patterns (section 3.3).

3.1 Design Rationales

The process of information propagation over social networks consists of three key factors: the information being spread, the people who spread the information from different places, and the processes and effects of the spreading [33]. In order to visually present these key factors and their relationships, we identify a number of design goals to be achieved in our system.

A. Visual narrative structure. To enhance visual pattern comparison, a visualization should encode key information into structurally organized component features which can be easily identified across multiple patterns. To facilitate an intuitive understanding about information diffusion, we design Whisper based on the metaphor of a sunflower and how its seeds are carried and dispersed, where the key components within Whisper visualization are visually analogous to the functional parts of a sunflower. The metaphorical pattern of a sunflower makes it easy for users to grasp the visual narrative of the information diffusion and enables an immediate comparison without spending much effort learning the meaning of each component.

B. Role identification. When tracing information diffusion in social media, it is crucial to identify key roles such as opinion leaders. It has been recognized that opinion leaders play decisive roles in the spread of information. According to Roger [41], "opinion leaders" refer to a small minority of people who bridge media and common users and determine the ways (positive/ negative, fast/slow, and far/near) in which information is propagated [33]. In contrast to opinion leaders, the majority of users, as studied by Romero et al. [42], act as passive information consumers. They do not forward the content to the network and are thus not actively engaged in information propagation. Therefore, representing opinion leaders and the information spread by them is crucial for tracing information diffusion in social media. Following this idea, tweets are visually classified into two categories, the tweets from media broadcasters and opinion leaders, and the tweets from other users.

C. Diffusion process unfolding. The multi-step retweeting behavior of users and opinion leaders should be simultaneously represented in order to capture the diffusion process within a given time window. The characteristics of a diffusion process have been identified: it follows a multi-step model [32] which states that an information flow first transits from mass to opinion leaders who, in the second step, significantly accelerate the spreading of the information. Through social media, the spreading of information can be accelerated in a very short period of time [34], and the propagation may travel in arbitrary directions in the social network simultaneously. Thus, the detailed retweeting behavior, "when did who retweet whom", needs to be illustrated along different directions simultaneously. Whisper seeks to represent such rich information through a collection of diffusion pathways on which users' retweeting behavior is shown at different levels of granularity. Each pathway is also a timeline whose time span is configurable to enable an exploration of the diffusion processes occurring between two chosen points in time.

D. Monitoring and scrutinizing. Visualization about a diffusion process should have both dynamic updating to support realtime monitoring and temporal exploration to support backward scrutinization of historical patterns. Whisper seeks to support two modes of visualization: (1) the dynamic view shows the tweets and retweets generated in real time, which helps capture what is happening about the topic of interest; (2) the static view with timeline control enables users to explore the historical data and to support in-depth visual analysis [40].

E. Visual clutter reduction. Reducing visual clutter is challenging when facing huge volume of social media data. As Whisper aims to provide a clear view of the spatiotemporal diffusion trend, any visual artifacts such as unnecessarily long traversal of tweets or edge crossings should be reduced by clutter reduction techniques [25]. Whisper seeks to effectively reduce visual clutter through novel layout methods. We use a flux line-drawing algorithm to create diffusion pathways that reduce crossings. Then, the layout of tweets and topic disc is further optimized to reduce the visual clutter at the global level. These layout methods will be described in detail later.

3.2 Visual Metaphors and Encodings

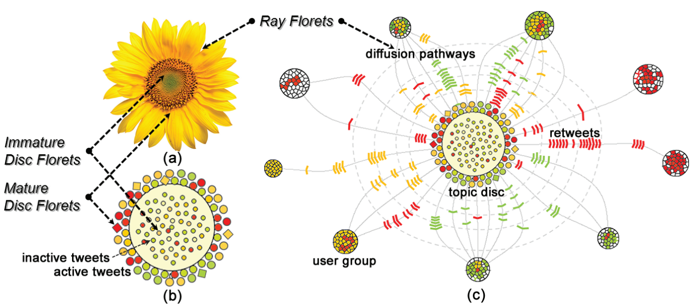

Fig. 2. Visualization design of Whisper: (a) the structure of a sunflower; (b) visual metaphor of disc florets; (c) visual metaphor of ray florets.

The overall design is based on a metaphor of a sunflower. The bloom of the sunflower is composed of thousands of disc florets inside a circular head, surrounded by ray florets, as shown in Fig. 2. The disc florets mature into seeds which are often dispersed away by wind, animals, or people. In analogy, the dots in the center in place of the sunflower disc florets represent tweets about topics of interests (Fig. 2(b)). The lines forming the shapes of the sunflower ray florets represent the diffusion pathways, tracing the path from the information source tweet to different groups of users who rebroadcast (retweet) the tweet. As in Fig. 2(c), user groups are represented by cluster icons at the end of the ray florets. This design that follows the design rationale **A** consists of three major components: topic disc, user group and diffusion pathway. We describe each of these components below.

Topic Disc. The topic disc is analogous to the central circular head of a sunflower. The immature disk florets are in the very center of the circular head, surrounded by mature disk florets. Metaphorically, as shown in Fig. 2(a) and 2(b), the dots packed in the center represent inactive tweets that have never been retweeted and the dots surrounding the topic disc are the active tweets that have been retweeted within a given period of time. A tweet is moved from the center to the outside ring when it gets retweeted for the first time. Tweets will fade out if they get no retweets for a period of time, making active tweets stand out from the large amounts of newly gathered tweet data.

User Group. To reveal how a post is being retweeted over different users (retweeters), we hierarchically group the retweeters based on their innate relations such as shared topic interests or geographic locations (design rationale **B**). Geographic grouping is provided by default considering its intuitiveness, computational efficiency and stability during real-time monitoring. The groups can be evenly laid out on a circle or placed into positions corresponding to their geo-locations (latitude or longitude coordinates). Examples of longitude and circular layouts can be seen in Fig. 1 and Fig. 10 respectively.

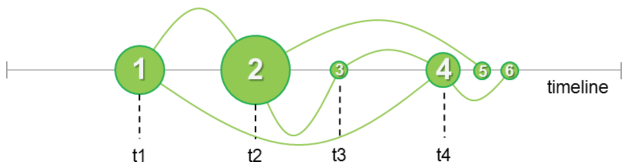

Fig. 3. The focused view of diffusion series on which the detailed retweeting behavior of users is visualized along a timeline.

Diffusion Pathway. The path linking a tweet to the retweet user groups illustrates a diffusion path along which the information of the tweet is diffused based on retweeting (design rationale **C**). As shown in Fig. 2, the area between the topic disc and the user group is mapped as a timeline between the starting and ending points of a user specified time period. When retweeting occurs within the given time period, a replica of the retweet is marked at the time point indicating when the retweeting happens. Diffusion pathways are traced as curved lines to reduce crossings and overlaps, which also mimic the shape of ray florets. Different diffusion patterns can be enhanced based on the curvature of the pathways to reveal information diffusion over diverse groups or within a few focused groups. The first case can be indicated by a set of widespread curved pathways, and the second case is indicated by a small and concentrated set of pathways.

When any of the pathways is focused upon, a "diffusion series" is shown to reveal the specific relationships among the users of the information source tweet and the succeeding retweet users (design rationale **B**-**D**). Fig. 3 illustrates the diffusion series. An information source tweet is posted by user 1, which gives the starting point of the diffusion series. User 2 retweets this tweet at time t_2, which is represented by an arc connecting two user marks at t_1 and t_2. Similarly, the arcs connecting the pairs of users 2 and 3, 2 and 5, 1 and 4, and so on, are occurring between different pairs of users at different time.

Visual Encoding. In Whisper, we encode information into the visual components consistently. **Color hue** is used to encode sentiments. Three pre-selected colors, red, orange, and green, are used to represent negative, neutral, and positive opinions respectively. For example, a green tweet glyph indicates the tweet message contains positive sentiment, and a user or a user group in green indicates the retweets from that user or group are positive on average within a given period. The sentiment of a tweet (positive, negative or neutral) is classified based on a sentiment score computed from the mean score of sentiment keywords in the tweet message, where the sentiment scores of keywords are given based on an affect dictionary [38][2]. The sentiment of a user group is computed based on the weighted mean scores of tweets that are retweeted by the group of users. **Opacity** is used to encode activeness. The activeness of a tweet is defined by how frequently it has been retweeted. The activeness of a user or a user group is defined by aggregating the activeness of all tweets by the user or by the user group. **Size** is used to encode expected influence. The expected influence of a tweet is given by the expected influence of the tweet user, which is defined by the number of followers the user has. Usually, users with more followers tend to be retweeted more frequently [14], but it also varies with context. The expected influence of a group is aggregated based on the influence of individual users assigned to the group. **Shape** is used to encode user types. A square tweet glyph indicates the tweet user has been recognized as a media outlet, journalist, or organization representative. A round glyph tweet indicates the user does not belong to any of these categories. Retweets are shown as arrow shape glyphs, with arrows pointing to user groups.

3.3 Interactions

Whisper incorporates a set of interactive functionalities that support further drill-down to data details along user, topic, spatial, or temporal dimensions.

Topic driven data exploration. Using Whisper, users can monitor a topic in real time by giving a keyword or a set of keywords. They can also query certain keywords to gather and explore historical tweet data. The keywords can be connected by logic operators such as "and" and "or" for more precise query. Both the monitoring and querying results can be stored in a local database, which enables data sharing.

Temporal exploration. Whisper supports temporal exploration in multiple ways. The timeline control allows users to slide back and forth to explore data within different time windows in the history. Users can also specify the time duration for retweets to be shown on both the diffusion pathways and on the diffusion series – a shorter time span allows a focus view on bursty topics.

Spatial zoom-in (hierarchical exploration). The capability of zooming into a user group enables users to explore the diffusion process within a location of interest, and to further see the activities of individual users within the group.

Highlighting. Elements in Whisper, including tweet/retweet glyphs and user group icons, can be focused on when the mouse hovers over the elements. When an element is focused on, a tooltip containing detailed information such as the tweet message is shown. In addition, elements that are connected to the focused element are also highlighted. We highlight the elements through color enhancement. For example, when a tweet is focused on, all its connected diffusion pathways and user groups are highlighted in brighter colors, whereas elements without direct connections to the focused one turn dim. The specific diffusion series of the focused tweet is also shown.

Filtering. Different types of data filters, including topic filters, influence filters, and sentiment filters, are provided in Whisper to support data reduction and facilitate subset comparison.

Switching direction of information flow. Although the visual design focuses on representing "where the information is propagated", it is also interesting to show "where the information came from." To incorporate the aspect of tracing information sources, a "convergence" mode is provided in addition to the original "diffusion" mode. In the "diffusion" mode, the retweets of a tweet are moved from the topic disc to the groups of users who retweet the tweet, while in the "convergence" mode, the retweets are moved from the group of users who post the original tweets toward the topic disc. Users can switch between the two modes to focus on either direction of information flow.

4 System Overview and Usage Scenario

As shown in Fig. 4, Whisper's system architecture consists of three primary components. First, in the raw data module, micro-blog data are gathered based on Twitter APIs[3]. Two data gathering modes, the streaming mode and the search mode, are supported by these APIs. The streaming mode provides a forward monitoring approach that allows users to monitor current and future events in a real time. In search mode, users can query the historical data. The spatial information of each tweet is also collected using the Google Geocoding API[4] while the data are gathered. These raw data are cleaned and stored in a

[2] The tweet sentiments are encoded into one of the three colors based on the signs of the computed sentiment scores.

[3] https://dev.twitter.com/docs

[4] When the tweets contain no geolocation/place tags, we use the location information provided in tweets and users' profiles as the locations of the tweets. Since the location field of a user and the referenced location in a tweet are unformatted in plain text and contains no coordinate information, we use Google Geocoding API (https://developers.google.com/maps/documentation/geocoding/) to translate

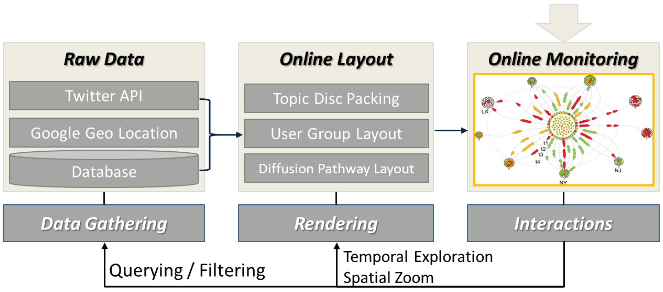

Fig. 4. The overview of Whisper visualization system.

database to support temporal explorations. The layout module supports efficient layout algorithms that transform and render the raw data immediately into visualizations in real time. In the whole process, no topic modeling except Twitter APIs are used. Finally, interactions are also supported. They feed back to the rendering and data module to enable online data exploration.

The user interface of the Whisper system consists of multiple components as illustrated in Fig. 1: 0) the Whisper visualization that displays the incoming tweet data; 1) the query keyword input box; 2) the layout selection drop down menu; 3) the sentiments filter; 4) the scrolling list that shows the incoming tweet messages based on the given topic (specified by the query); 5) the topic filter; 6) the influence filter (to select tweets based on the users' expected influence) and keyword filter (to display tweets that contain only the selected keyword(s) when multiple keywords are given in the query box)[5]; 7) the interactive timeline that summarizes the volume of incoming tweets and the average sentiments within each time interval; and 8) the diffusion series on which the detailed retweeting behaviors of users are visualized.

Consider the following scenario for using the Whisper system. Jean is a political columnist. One of her primary jobs is to trace important political news and events. She uses Whisper to help her on these tasks. Recently, she is monitoring the US presidential primaries of the Republican Party. She traces candidates' names in Whisper. Whisper initially provides a worldwide, global view of the related tweets. Jean zooms into the USA region and explores the collective responses from various states. She filters the tweets by candidate names and compares the differences in the collective responses of these candidates. Interestingly, she finds that some tweets propagated by some primary media such as CNN arouses a significant impact in public. Whisper visualizes the dramatic spreading of these tweets across various states along the parallel diffusion pathways. When Jean focuses on a single pathway, a diffusion series is shown to illustrate the detailed diffusion patterns of "when did who retweet whom". Tooltips allow her to see the content and user of each tweet when she uses mouse to hover over these objects. Again, Jean finds that many tweets become popular after being retweeted by an important opinion leader. Thus, she decides to follow him/her on Twitter and also plans to contact him/her for an interview. Jean saves the data in a file and sends it to her colleagues for further discussions by e-mail.

5 Implementation

In this section we describe the methods used to implement each of the key design components, including topic disc (section 5.1), diffusion pathways (section 5.2), and user groups (section 5.3), as well as methods that are used to reduce visual clutter and assemble all these components together (section 5.4).

5.1 Layout of Topic Disc

Topic disc, located at the center of Whisper display, is to carry the information sources, including active tweets (tweets that get retweeted within a given period of time) and inactive tweets (tweets that have no retweets). The space can be dense depending on the volume of incoming tweets. Hence packing large amounts of tweets aesthetically and efficiently into the topic disc is crucial.

Inactive tweets and active tweets are laid out in different ways. The inactive tweets are placed inside the topic disc using the well-known sunflower packing algorithm proposed by Vogel [48]. In our implementation, we sort inactive posts in an increasing order based on their influences as defined in section 3.2. The unimportant small posts are hence laid out in the center whereas the important bigger ones are laid out on the edge to mimic the center disc of a real sunflower.

The active tweets are placed as a ring surrounding the topic disc. When there are considerable amount of active tweets coming at the same time, the space grows into multiple concentric rings. An example can be seen in Fig. 2(c). The outmost rings contain the latest tweets, whereas those inner rings contain relatively older tweets in a temporal order.

5.2 Layout of Diffusion Pathways

A diffusion pathway, as described in section 3, is meant to show the concurrent retweeting activity and simultaneously highlight the trends of information propagation. These are conflicting requirements because massive concurrent retweeting activity can easily lead to visual clutter and trends are not automatically revealed if the activities are displayed literally. The design criteria cannot easily be fulfilled by existing edge routing techniques. In this section, we introduce a novel edge routing algorithm based on an "information diffusion field" model.

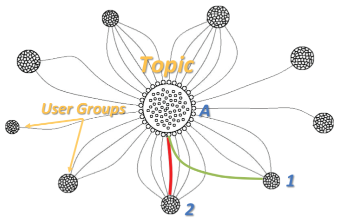

Fig. 5. The diffusion field lies between a single topic and multiple user groups, with flux line curves coincident with the shape of a sunflower. The topic node adheres to positive charges and are represented as a sunflower disc. User group nodes adhere to negative charges and is represented as voronoi icons.

Diffusion field and flux lines. Our algorithm is inspired by the properties of electric flux lines: 1) flux lines in an electric field never cross each other since they have a repelling effect upon their neighboring flux lines; 2) flux lines have force tensions which make them as short as possible. Unlike traditional edge routing algorithms such as hierarchical edge bundles [30] that bundles lines together, the two properties of the flux-line drawing make it a perfect choice in our case. First, it enables tracing information diffusion processes triggered by different messages since the repelling effect reduces visual clutter and provides additional space between lines. Second, the shorter lines enable users to see diffusion trend without tracing long curves [49]. Furthermore, flux line curves are coincident with the shape of the ray florets in our visual metaphor of a sunflower.

We introduce an artificial *diffusion field* model to mimic electric fields and to generate flux lines. The model is defined based on a bipartite graph that consists of two types of nodes: topic nodes and user group nodes. Topic nodes are treated as positively charged particles whereas user group nodes are negatively charged particles. A 2D diffusion field is generated once a bipartite graph is laid out on the surface. Fig. 5 gives an example of diffusion field based on a single topic graph. When a positive charge is assigned to an active tweet in the diffusion field, the tweet is forced to move from a topic node to a user group along a flux line in analog to an information diffusion process.

Drawing flux lines. In physics, the electric field, $\overline{E_i}(x,y)$, of a point charge q_i at a given position (x,y) on a 2D surface is defined by the electric force, $\overline{F}(x,y)$, on a testing charge q located at (x,y):

them into standard geolocation format.

[5]Due to space limitation users can see up to five keywords on the screen. A button is added in the end of the list when there are more than five keywords. When users click on the button, more keywords will be shown in a popup window in which users can make a selection.

$$\overline{E_i}(x,y) = \frac{\overline{F}(x,y)}{q} = k_e \frac{qq_i}{qr^2}\overline{r_i} = k_e \frac{q_i}{r^2}\overline{r} \quad (1)$$

where r_i is the distance between charge q and q_i; $\overline{r}$ is a normal vector that points from q_i to q; k_e is a physical constant. Furthermore, the total electric field $\overline{E}$ generated by a collection of point sources is the vector sum of the electric fields of the individual charges:

$$\overline{E}(x,y) = \sum_i E_i = \sum_i k_e \frac{q_i}{r_i^2}\overline{r} \quad (2)$$

A flux line $\rho(t)$ of the field $\overline{E}$ is a curve that is tangent to $\overline{E}$ everywhere along the line, that is:

$$\frac{\partial \rho(t)}{\partial t} = \overline{E}(\rho(t)) = \overline{E}(x,y) \quad (3)$$

where $\rho(t)$ is a parametric function of the flux line and t is the parameter. With a specific t, $\rho(t)$ gives a certain point on the flux line, e.g. $\rho(0) = (x_0, y_0)$ indicates the starting point of the flux line which provides the initial condition of the above differential equation. This parametric function $\rho(t)$ provides a way of drawing flux lines in fields. It can be computed by integrating differential equation (3) incrementally in a discrete domain:

$$\rho(t+\Delta t) = \rho(t) + \int_t^{t+\Delta t} \overline{E}(\rho(t))dt \quad (4)$$

The above equation can be solved by a fourth-order Runge-Kutta integrator [18].

The standard computation of flux lines, although have no crossings, may lead to semantic confusion when these flux lines are used to represent diffusion flows. One example can be seen in Fig. 5. Suppose the green curve is a diffusion flow that illustrates the diffusion of tweets from topic A to user group 1. A standard field model may force the line to connect topic A to group 2, as shown by the red curve. This occurs when group 2 is sufficiently close to topic 1 and thus has larger attraction force to the flux than the force from group 1. In this case, some physical laws need to be violated in order to allow necessary flux line crossings in order to generate a result with correct semantics. Specifically, we adaptively change the magnitude of the source charges (topics and user groups) in the diffusion field for each pathway. Using the above example again, when laying out the pathway from topic A to group 1, we significantly increase the charge magnitude of group 1 and decrease the charge magnitudes of other groups. Well selected increasing and decreasing ratios provide esthetic and correct results. In the case of the sunflower diffusion field as shown in Fig. 5, we set this ratio to 30.

Activity embedded in diffusion field. Once the flux lines are laid out, we can display the detailed diffusion process by embedding retweeting activity in temporal order along the line. The shortest distance of the entire flux line is proportional to the time duration specified by users or by a system default. A flux line connects the information source (the original tweet) in one end, and the retweet users group in another. With the given time duration, all the retweets of the original tweets, if flowing to the same user group, will be shown as retweet glyphs on this flux line in temporal order, with distance to center indicating the time of retweeting.

5.3 Layout of User Group

Users inside a user group (e.g. a geo-community) can be laid out similarly as tweets in the topic disc. Our system adaptively chooses the efficient sunflower packing algorithm to lay out users in a user group for real-time monitoring and switches to a voronoi layout method [19] to facilitate static comparison.

User groups are placed in an outer ring based on two layout options: (1) longitude layout: projects user groups on the circle based on the retweet users' location longitude available from the twitter APIs or obtained from the Google Geocoding API; (2) circular layout: projects the user groups onto a circle by equally dividing the circle arc. In the first case, a world map [6] is shown as the background to provide additional visual cues for the groups' geographic locations.

[6]The world map is projected from the north pole based on lambert azimuthal equal-area projection (http://en.wikipedia.org/wiki/Lambert_azimuthal_equal-area_projection).

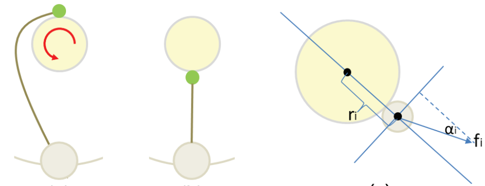

Fig. 6. Minimize the curve length based on a force model. a) Direct connection without optimization. b) Optimized connection. c) Parameters used in the force model.

5.4 Integration and Refinement

The last step is to integrate all visual components together. We use a radial layout scheme in which the topic disc is put at the center of the diffusion field, surrounded by a collection of user groups. To further reduce line crossings, we incorporate two strategies: reordering the positions of active tweets, and rotating the topic disc.

Reordering. We reorder the positions of active tweets when the user groups are arranged based on the circular layout. We construct a bipartite graph $G(\{V_A, V_B\}, E)$ where V_A and V_B are the vertex sets that contain the active tweets and the user groups respectively; E is the edge set consisting of all the diffusion pathways connecting nodes in V_A to nodes in V_B.

Intuitively, the reordering can be considered as a process in which we first anchor the positions of user groups and then use forces along the edges to drag and move the active tweets along the edge of the topic disc. The order is optimized when the force tension of the edges is minimized.

Specifically, we first project the positioning of all user groups onto an one-dimensional ordinal space according to their layout orders. In this reordering process, the positions of user groups remain unchanged while the positions of active tweets are reordered using the following stress model given the bipartite graph G:

$$min \sum_{i,j} \omega_{ij} ||x_i - y_j||^2 \quad (5)$$

where x_i and y_j are one-dimensional coordinates of the i-th active tweet $a_i \in V_A$ and the j-th user group $b_j \in V_B$, ω_{ij} is the edge weight betweet a_i and b_j, defined by $1/d_{ij}$, the inverse distance betwee a_i and b_j if $e_{ij} \in E$, and 0 otherwise. The optimal solution of the coordinates can be found by using spectral relaxation [54]. To aovid dramatic changes and keep users' mental map, the reordering is only performed when the overall stress is over a predefined threshold.

Rotating the topic disc. After the optimal ordering of user groups is determined, we further rotate the topic disc to reduce line crossing. The rotation can be determined by considering an elastic band connecting the topic disc with the fixed surrounding user groups [13]. As illustrated in Fig. 6(a)(b), a long curved band will force the topic disc to rotate in order to minimize the stress. We use a spring force model defined as follows:

$$min \sum_i (f_i \times r_i \times sin(\alpha_i)) \quad (6)$$

where, as illustrated in Fig 6(c), f_i is the spring force along the i-th edge, r is the distance from an active tweet to the center of the topic disc. α_i is the orientation of the edge. This model rotates the topic disc to minimize the overall force tensions and results in an angle that makes the distances between all user groups and the associated active tweets optimally balanced. Furthermore, we reduce the edge intensity according to the curvature of the pathways so that the visual attention will not be biased to the opposite directions.

5.5 Time complexity

We discuss the time complexity of each step described above. The layouts of the topic disc and user groups are based on the sunflower packing algorithm which efficiently creates phyllotaxises [37] in $O(n)$ time where n is the number of tweets or users that need to be packed. A single diffusion pathway with m discrete sample points can be computed in $O(m)$ time by the Runge-Kutta integrator [18]. Both of the visual clutter reduction methods discussed in Section 5.4 has the time complexity of $O(uv)$ for each iteration where u is the amount of pathways or active tweets and v is the amount of user groups. In most cases, u and v are small numbers within a small time window. Usually, the force models can quickly converge within limited iterations k when the parameters are carefully selected. For example, in our implementation, we set the magnitude of the unit spring force as 6E-9 and the energy threshold as 1E-5, which lead to 32 iterations for convergence on average. Therefore, the overall time complexity is $O(kuv+n+um)$ which is efficient for real-time scenarios (see section 6.1 for performance evaluation).

6 Evaluation and Discussion

In this section, we first demonstrate the performance of our system. After that we show the usefulness of our visualization with current events from twitter data, including natural emergency and political campaigns. Finally we present feedback from domain experts and discuss the insights provided by our design.

6.1 Performance Evaluation

We tested the performance of Whisper system based on the incoming tweet flows against different TPS (tweet per second). Our experiments were conducted on a Mac Pro (2.4 GHz Intel Core i5, 4GB Memory, and Mac OS X Lion 10.7.4). A dataset was collected before the evaluation. It contains 19,300 tweets and 38.4% of which are retweets. The tweet flow's average TPS is 5 (STD: 3.04). In the experiments, we simulated a 5-minute monitoring scenario with respect to different TPS. We used the real tweet data repeatedly but modified the timestamps of the tweets in order to generate a tweet flow under a given TPS. The new data strictly follow original data distributions both in locations and in retweet portions. We tested the TPS from 10 (10 TPS $\cdot$ 60 sec $\cdot$ 5 mins = 3000 tweets within 5 mins) to 5000 (1,500,000 tweets within 5 mins). The sliding time window size was set to 3 minutes and the testing was based on the longitude layout. The layout and rendering time costs are shown in Fig. 7. The results show that, when TPS is less than 1000, Whisper is able to layout and render the data in real time within a 3-minute time window given the total time cost is less than 1 second. This suggests that the visualization performance of Whisper is efficient in many real-time scenarios. However, in case of online monitoring, the Internet connections, bandwidth and Twitter API access rate may cause significant delay for displaying the data, which can be hardly tested.

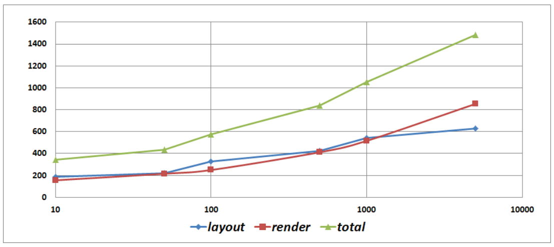

Fig. 7. Performance evaluation of Whisper system. X-axis is the tweet per second (TPS) and Y-axis is the time cost in millions seconds.

6.2 Case Studies

We present two examples, a recent earthquake in Japan and political campaigns in US to illustrate how Whisper could be used to monitor real-time events.

Earthquake. We take the recent earthquake as an example to illustrate how Whisper facilitates the exploration of people's reactions and the temporal and spatial diffusion patterns. On March 15, 2012 a 6.8 magnitude earthquake shocked the northern coast of Hokkaido island. It triggers a series of aftershocks and tsunami waves. The event caught global attention because the location was one of the areas in northeast Japan devastated by the magnitude 9.0 earthquake and tsunami last year. We monitored this event based on the keyword "earthquake". As shown in Fig. 1, the news was dispersed over the globe, and people from Japan paid the most attention. While this is a natural emergency, the information was quickly spread with negative sentiments as the majority (Fig. 8(a)). For example, people from other countries such as Australia were initially concerned about the Pacific-wide tsunami threat, and some from the Philippines were relieved when hearing about the safety of a celebrity (see Fig. 8(b) for the tweet about Greyson Chance).

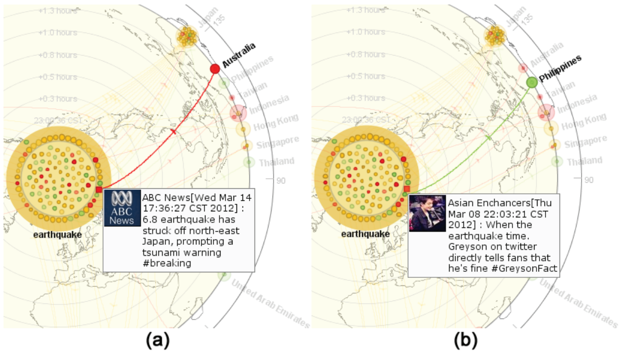

Fig. 8. The information about an earthquake was spread with different sentiments. (a) People from other countries such as Australia were initially concerned about the Pacific-wide tsunami threat. (b) Some from Philippines were relieved when hearing about the safety of a celebrity.

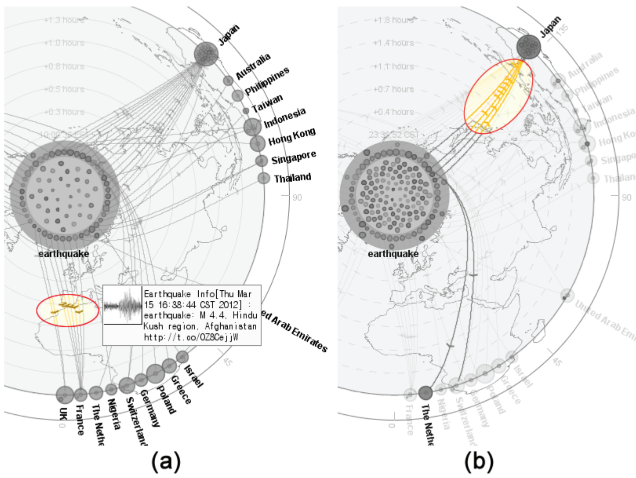

Fig. 9. The spatial and temporal spread patterns of different pieces of information were revealed in Whisper. (a) The burst activity was indicated when the retweet glyphs are aligned on the same ring. (b) The tweet about the earlier major earthquake continued to get a sequence of retweets from Japan.

Whisper is also able to uncover the spatial and temporal spreading patterns of different pieces of information. In Fig. 9(a), a tweet regarding an aftershock happening in the night was instantly retweeted by a number of users. This burst of activity is indicated when the retweet glyphs are aligned on the same ring. The tweet about the earlier major earthquakes continued to get a sequence of retweets from Japan (Fig. 9(b)), but the same information written in English was retweeted more broadly from other geo-communities (Fig. 1).

Political campaigns. We have closely tracked the recent hot discussions on Twitter regarding the US presidential primary elections of the Republican Party. This example is used to demonstrate how Whisper can effectively disclose community response patterns on an interesting topic.

The voting results came out on so-called "Super Tuesday", March 6 this year. A set of keywords including the names of the candidates, e.g., "mitt romney", "romney, gingrich, santorum, paul", were used to track the related activity. Through summarizing the community sentiments, Whisper was able to capture the disparity of the collective response in different geo-communities during the night. As shown

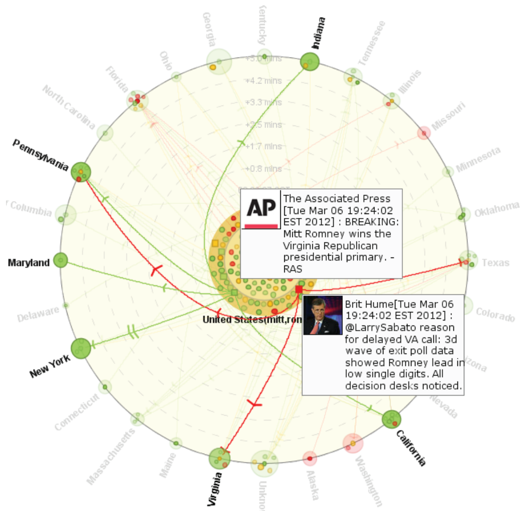

Fig. 10. Whisper captures the disparity of collective response in different geo-groups during Super Tuesday night (for US presidential primaries). The diffusion pathways highlighted in green show the retweets about Romney's victory, whereas the diffusion pathways highlighted in red show the retweets about Romney's problem.

in Fig. 10, users from several states were actively re-broadcasting the news about candidate Mitt Romney's victory in Vermont (pathways in green). At the same time, some users from Louisiana and California retweeted the comments made by senior journalist Howard Fineman about "Romney's problem" (pathways in red).

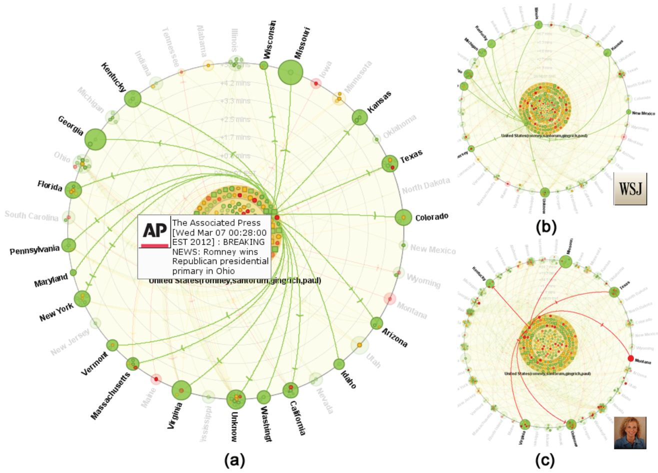

Fig. 11. The spatial diffusion patterns regarding the Republican presidential primaries and caucus results on Super Tuesday, triggered by media outlets, such as the Associated Press (a), Wall Street Journal (b), and an opinion leader (Congresswoman Schultz) (c).

The role of media outlets and opinion leaders in driving the user activities in social media can also be unfolded in Whisper. Fig. 11 shows the diffusions triggered by two media outlets, the Associated Press (AP) and the Wall Street Journal (WSJ), and by an opinion leader, the Congresswoman D. Wasserman Schultz. AP's announcement of Romney's victory was more extensively retweeted from most of the states in US, compared with a similar message from WSJ[7]. Congresswoman Schultz is a Democratic representative of Florida. Her comment about Romney on the night got retweeted from several states including Missouri, Texas, Montana and Kentucky, which interestingly exemplified how today's politicians extend influence across congressional districts through social media. In Whisper, users can differentiate opinion leaders and media outlets from ordinary Twitter users based on the sizes and shapes of the tweet glyphs (the sizes encode the expected influence of the tweet users and the shapes are drawn based on a predefined list of media outlets).

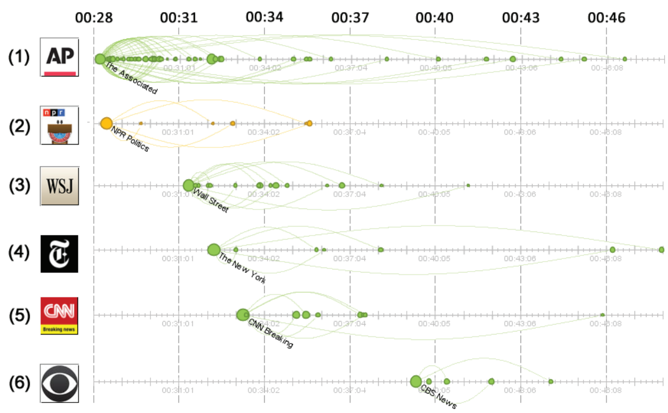

Fig. 12. Temporal patterns of diffusions triggered by a number of media outlets. AP is the first media that reports Romney's victory and received the most retweets then all other media.

The temporal patterns of diffusion can be further discovered in diffusion series[8]. As shown in Fig. 12, AP was the first media that reported Romney's victory. A number of other media outlets (e.g., NPR, The New York Times, CNN, etc.) also reported the same event in about 10 minutes. Several interesting observations can be drawn from this figure. First, an outlet's immediateness of reporting an event critically determines the number of retweets it would receive – the first outlet AP attracted the most attention than all the other outlets. Second, the message content is also an important factor – for example, NPR reported the event almost at the same time as AP but attracted much less attention since the tweet only provided information about polling rate without conclusive description. In contrast, other media that clearly reported that Romney was the winner received more retweets even they reported later than NPR. In addition, the diffusion instances all followed a similar pattern in which the spreading significantly accelerated within the first few minuets and then slowed down.

6.3 Interview with Domain Experts

To evaluate the quality of our design, we conducted in-depth interviews with three experts. Due to the interdisciplinary nature of this research, we chose experts from three relevant professional areas. The first expert is a political scientist with expertise in social network and social influence. The second expert is a communication Ph.D. with expertise in social networks and epistemology. The last expert is a design expert currently working for a social TV analytics company.

Interview methodology. In each of the interview sessions, we started with a tutorial to explain the purpose and features of Whisper. We then asked the experts to use Whisper to monitor any ongoing topics or events. We also asked them to explore the pre-recorded twitter data about several large-scale events. After they fully explored the tool's capabilities, we conducted a semi-structured interview guided by the following questions:

1. How does the topic become popular?
2. When and where does the topic attract attention? By which community?
3. Is the topic of local or global interest?
4. How do different communities respond to the topic?
5. Are there any media broadcasters or opinion leaders discussing this topic? When and how does that happen?
6. How does the tool help distinguish media broadcasters, opinion leaders or normal users involved in the discussion of the topics?

[7]Based on our rendering algorithm, a tweet will be positioned as close as possible to the initial destination group(s) in order to shorten the expected pathways, so pathways resulted from opposite directions reflect that there is a progression on spatial diffusion toward different directions.

[8]When users click on an active tweet (not however on), the related diffusion series will be selected in the comparison view within a pop-up window.

7. How does information/news propagate differently over time for different topics? How does information/news propagate differently over locations/communities for different topics?

Besides these questions, we asked the experts to elaborate their thoughts based on their domain knowledge. Each of the interviews last approximately 1.5 - 2 hours. We recorded the experts' responses in detail and took notes of their comments.

Results. We summarized the domain experts' feedback as follows, where the three experts are denoted as User P (political science), C (communication), and D (design), respectively.

Overall visual and interaction design. All three users were impressed by the real-time monitoring and visualization when they realized the demo program was showing data in real time. User C said "Wow! It's amazing" when he saw the monitoring feature. He also commented, "the visual vocabulary in this design is powerful; it provides many levels of detail and flexibility to generate and test hypotheses about an event." User D commented, "I like this visualization! [Compared with tree-like diffusions such as Google Ripples,] in your visualization, the grouping of users creates layers and you get space to show the spatial and temporal interaction." Particularly, all three users recommended the *community zoom-in feature*. User D commented, "users are grouped together as communities at first, but you can still see people inside the communities; the first level simplification allows you to introduce another level of complexity [referring to the spatial-temporal interaction]."

Topic trend discovery. All the users agreed that it is easy to see how a topic becomes popular. User C commented, "I can see this from the ratio of tweet-retweet density at the center and outer ring." User D said, "I see three [visual] popularity measures: how many tweets, how many retweets, how many geo-communities are influenced."

Community response. Users acknowledged that the summarization of community sentiment (derived from the original tweets) makes both the narrative and visual presentation more interesting. User P commented, "I can have a sense of what sentiment is from what place." User P especially appreciated the sub-topic filtering feature which then can be used to compare the community sentiments toward different subjects or politicians when a given topic involves multiple subjects. He commented, "I can see the polarized sentiments about Santorum."

Media broadcasters and opinion leaders. User C acknowledged the advantages of differentiating media broadcasters from the rest of the users, but using different glyph shapes may be too subtle to distinguish different types of users, especially when the tweet density is high. On the other hand, according to User C, the retweet series showing the exact timing of retweets is useful for examining an event.

Our users also highlighted many potential applications of Whisper. All three users acknowledged that the current design has great potential, especially in news consumption and marketing analysis. User P suggested, "it would be useful to incorporate the location context such as per capita GDP or population density in this visualization so that the analysts would see the real impact of the places." User C commented, "a journalist would like to report a topic which has not got much attention in one place but is becoming popular elsewhere; your tool has the potential to do this."

Discussions. Despite the above positive feedback, our users also had some concerns after they used Whisper for the first time that are worth mentioning:

- Plentiful information at a first glance. Both users P and C felt that the overall information display, although impressive, was overwhelming when they first saw it.
- Space/time confusion. User C felt that in the circular layout, the retweet traces on the temporal rings resulted from the spatial-temporal dynamic layout can be confusing, and spatial patterns are difficult to compare across frames, especially when there is huge amount of information quickly come in.
- Counter-intuitive time progression. Both users C and D felt that the retweets moving toward the geo-locations as time progresses was counter-intuitive because the latest retweets were placed farthest from the retweeters' locations and gradually moved closer.

The first concern comes from the complexity of the data. The data gathered in real time are indeed overwhelming and it is hard to capture from raw data the potentially interesting elements (e.g., interesting tweets or influential users). A design decision made in Whisper is to visually guide users to spot potentially interesting elements from massive diffusion instances, with an overview picture of the dynamics of diffusion over time. Unlike traditional diffusion visualization, our design can support further interpretation and in-depth analysis about the captured patterns involving information content, people, time, space, as well as the complex interplay among them. Indeed, as one of the experts commented, a key contribution of our design is its capabilities to generate new hypotheses and support in-depth analysis of information diffusion about a new event. When the displayed information increases due to incoming data, we shall guide users to utilize all the capability built in Whisper, including filters and hierarchical exploration, to construct a focused representation of massive incoming information.

The second issue about space/time confusion only appears in the circular layout and is also a result of the design decision. We believe that the reduction of visual clutter (edge crossing) has a huge impact on the ability to see patterns. Hence the dynamic layout, such as the reordering user groups described in section 5.2, was introduced to enable a clear overview of the data. On the other hand, we also provide an alternative "longitude" layout for users to compare patterns over space if they are interested.

The third issue interestingly presents the diverse, interdisciplinary perspectives on understanding information diffusion – it can be understood as a social process of people consuming information, or as epidemic spreading over a population. In our design we take the second perspective, where information spreads from information sources to users (who retweet the information) and to the users' locations, as time progresses. During the interview, we found users have turned to appreciate such alternative perspective. Hence we believe this alternative perspective is compatible and also helpful in directing users' focus to the information being spread.

Overall, Whisper is appreciated by the three experts with different backgrounds. The interview also helped sort out the major strengths and the next steps of the current design.

7 Conclusion

In this paper, we presented Whisper, an interactive visualization for monitoring information diffusion of micro-blog data in real time. It incorporates a novel design in which we encode multidimensional information about a diffusion process into different functional parts of a sunflower. Particularly, the mature/immature disc florets are used to represent active/inactive tweets about a given topic. The ray florets are used to represent diffusion pathways that portray how the information diffused from the information sources at center to the surrounding user groups. This design intuitively represents the collective response, multi-step information flow as well as temporal heterogeneity of typical information diffusion processes. Our evaluation, including case studies and in-depth interviews of three domain experts from various related areas, demonstrates the usefulness of Whisper and also verifies our primary design goals.

For future work, we plan to develop a web version of Whisper to gain feedback from average users. An interesting extension is to incorporate other forms of communication carried out in social media, such as the "in-reply-to" communication in Twitter. We also plan to incorporate visual encodings of external contexts about the elements in the diffusion process in order to support experts' additional needs for investigating the relationship between information consumption and a broader context such as economics.

Acknowledgments

We thank all the domain experts for their contributions to the system evaluation and the anonymous reviewers for their valuable comments. This work was supported in part by grant HK RGC GRF 619309 and a grant from Microsoft Research Asia. We also gratefully acknowledge the support of the LazerLab at Northeastern University, supported by MURI grant #G00003585 and IARPA/HHRL #599437.

REFERENCES

[1] http://www.neoformix.com/Projects/TwitterStreamGraphs.

[2] monitter. http://monitter.com/.

[3] Trendistic. http://trendistic.indextank.com/.

[4] Tweetstats. http://tweetstats.com/trends.

[5] twinitor. http://www.twinitor.com/.

[6] Wefollow. http://wefollow.com/.

[7] Osama Takedown Sets Twitter Record, 2 May, 2011. Wired Magazine. http://www.wired.com/epicenter/2011/05/osama-take-down-sets-twitter-record/.

[8] NYTLabs Cascade: How Information Propagates through Social Media, 22 April 2011. Visualization Blog on Information Aesthetics.

[9] Pulse of the Nation: U.S. Mood throughout the Day Inferred from Twitter, 23 July 2010. Visualization Blog on Information Aesthetics.

[10] Google+ Ripples: Revealing How Posts are Shared over Time, 31 October 2011. Visualization Blog on Information Aesthetics.

[11] Riot Rumours : How Misinformation Spread on Twitter During a Time of Crisis, 8 December 2011. Visualization Blog on Information Aesthetics.

[12] D. Archambault, D. Greene, P. Cunningham, and N. Hurley. Themecrowds: Multiresolution summaries of twitter usage. In *Proceedings of the 3rd International Workshop on Search and Mining User Generated Contents, ACM CIKM*, pages 77–84, 2011.

[13] D. Archambault, T. Munzner, and D. Auber. Topolayout: Multi-level graph layout by topological features. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):305–317, 2005.

[14] E. Bakshy, J. Hofman, W. Mason, and D. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Fourth International Conference on Web Search and Data Mining*, 2011.

[15] A. Brew, D. Greene, D. Archambault, and P. Cunningham. Deriving insights from national happiness indices. In *IEEE ICDM SENTIRE Sentiment Elicitation from Natural Text for Information Retrieval and Extraction*, pages 53–60, 2011.

[16] L. Brown. *Innovation diffusion: A new perspective*. Methuen (London and New York), 1981.

[17] M. Burch and S. Diehl. Timeradartrees: Visualizing dynamic compound digraphs. In *Computer Graphics Forum*, volume 27, pages 823–830, 2008.

[18] J. Butcher and J. Wiley. *Numerical methods for ordinary differential equations*, volume 2. Wiley Online Library, 2003.

[19] N. Cao, D. Gotz, J. Sun, and H. Qu. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2581–2590, 2011.

[20] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[21] E. Cuvelier and M. Aufaure. A buzz and e-reputation monitoring tool for twitter based on galois lattices. In *Proceedings of the International Conference on Conceptual Structures for Discovering Knowledge*, pages 91–103, 2011.

[22] M. De Choudhury, Y. Lin, H. Sundaram, K. Candan, L. Xie, and A. Kelliher. How does the data sampling strategy impact the discovery of information diffusion in social media. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pages 34–41, 2010.

[23] M. Dork, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1129–1138, 2010.

[24] D. Dorling, A. Barford, and M. Newman. Worldmapper: the world as you've never seen it before. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):757–764, 2006.

[25] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.

[26] S. Few. *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc., 2006.

[27] M. Granovetter. Threshold models of collective behavior1. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[28] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Vizualization*, pages 115–123, 2000.

[29] C. Ho, C. Li, and S. Lin. Modeling and visualizing information propagation in a micro-blogging platform. In *International Conference on Advances in Social Network Analysis and Mining*, pages 328–335, 2011.

[30] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.

[31] P. Kallas. 48 Free Social Media Monitoring Tools. http://www.dreamgrow.com/48-free-social-media-monitoring-tools/.

[32] E. Katz. The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 21(1):61, 1957.

[33] E. Katz and P. Lazarsfeld. *Personal influence: The part played by people in the flow of mass communications*. Transaction Pub, 2006.

[34] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the ACM International Conference on World wide web*, pages 591–600, 2010.

[35] Y. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. Tseng. Analyzing communities and their evolutions in dynamic social networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(2):8, 2009.

[36] A. Marcus, M. Bernstein, O. Badar, D. Karger, S. Madden, and R. Miller. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the ACM Annual Conference on Human factors in Computing Systems*, pages 227–236, 2011.

[37] P. Neumann, S. Carpendale, and A. Agarawala. Phyllotrees: Phyllotactic patterns for tree layout. In *IEEE VGTC Symposium on Visualization Joint Eurographics, EuroVis*, pages 59–66. The Eurographics Association, 2006.

[38] F. Å. Nielsen. Afinn, informatics and mathematical modelling, technical university of denmark, http://www2.imm.dtu.dk/pubdb/p.php?6010, 2011.

[39] M. Ogawa and K. Ma. Stargate: A unified, interactive visualization of software projects. In *IEEE Symposium on Pacific Visualization*, pages 191–198, 2008.

[40] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1325–1332, 2008.

[41] E. Rogers. *Diffusion of innovations*. Free Press, 1995.

[42] D. Romero, W. Galuba, S. Asur, and B. Huberman. Influence and passivity in social media. *Machine Learning and Knowledge Discovery in Databases*, 6913:18–33, 2011.

[43] D. 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 ACM International Conference on World Wide Web*, pages 695–704, 2011.

[44] C. Schmid. *Statistical graphics: design principles and practices*. Krieger Pub Co, 1992.

[45] J. Schubring. 12 Social Media Monitoring Tools Reviewed. http://sixrevisions.com/tools/12-social-media-monitoring-tools-reviewed/.

[46] B. Shneiderman. Tree visualization with treemaps: 2d space-filling approach. *ACM Transactions on Graphics*, 11(1):92–99, 1992.

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

[48] H. Vogel. A better way to construct the sunflower head. *Mathematical Biosciences*, 44(3-4):179–189, 1979.

[49] M. Wertheimer. *Gestalt theory*. Hayes Barton Press, 1900.

[50] S. Wu, J. Hofman, W. Mason, and D. Watts. Who says what to whom on twitter. In *Proceedings of the ACM International Conference on World Wide Web*, pages 705–714, 2011.

[51] J. Yang and S. Counts. Comparing information diffusion structure in weblogs and microblogs. In *Proceeding of the International AAAI Conference on Weblogs and Social Media*, 2010.

[52] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. In *Proceeding of the International AAAI Conference on Weblogs and Social Media*, pages 355–358, 2010.

[53] S. Ye and S. Wu. Measuring message propagation and social influence on twitter.com. In *Proceedings of the International Conference on Social Informatics*, pages 216–231, 2010.

[54] H. Zha, X. He, C. Ding, M. Gu, and H. Simon. Spectral relaxation for k-means clustering. *Advances in Neural Information Processing Systems*, 2:1057–1064, 2002.