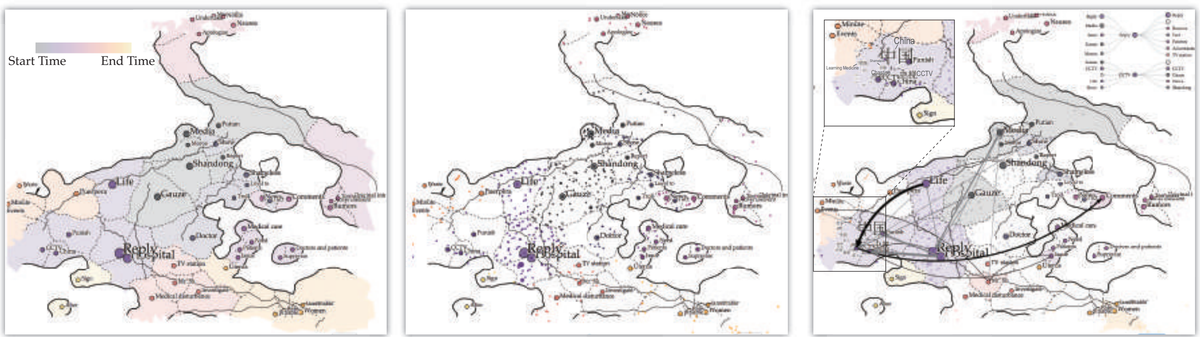
E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media

Siming Chen[1] Shuai Chen[1] Lijing Lin[1] Xiaoru Yuan[1*] Jie Liang[2†] Xiaolong Zhang[3‡]

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University, China
2) Faculty of Engineer and Information Technology, The University of Technology, Sydney, Australia
3) College of Information Sciences and Technology, Pennsylvania State University, USA



(a) (b) (c)

Figure 1: In E-Map constructed from the unstructured social media messages, a city is represented by a keyword extracted from events, where color gradient conveys the temporal evolution of events (a). The towns surrounding the city are shaped by the messages of specific keywords. Rivers represent highlighted social media users' behaviors of reposting (e.g. retweeting) (b). Users' trajectories and connections on the map encode the user behaviors of discussing different themes (curved black trajectories) and the information diffusion directions (straight gray connections). E-Map supports multi-level spatial and temporal exploration (c).

Abstract

Significant events are often discussed and spread through social media, involving many people. Reposting activities and opinions expressed in social media offer good opportunities to understand the evolution of events. However, the dynamics of reposting activities and the diversity of user comments pose challenges to understand event-related social media data. We propose E-Map, a visual analytics approach that uses map-like visualization tools to help multi-faceted analysis of social media data on a significant event and in-depth understanding of the development of the event. E-Map transforms extracted keywords, messages, and reposting behaviors into map features such as cities, towns, and rivers to build a structured and semantic space for users to explore. It also visualizes complex posting and reposting behaviors as simple trajectories and connections that can be easily followed. By supporting multi-level spatial temporal exploration, E-Map helps to reveal the patterns of event development and key players in an event, disclosing the ways they shape and affect the development of the event. Two cases analysing real-world events confirm the capacities of E-Map in facilitating the analysis of event evolution with social media data.

Keywords: Social Media, Event Analysis, Map-like Visual Metaphor, Spatial Temporal Visual Analytics

*e-mail: {siming.chen, shuai.chen, lijing.lin, xiaoru.yuan} @pku.edu.cn. Xiaoru Yuan is the corresponding author.
†e-mail: christy.jie@gmail.com
‡e-mail: lzhang@ist.psu.edu

1 Introduction

Social media plays an important role in discussing and disseminating information about significant events. Nowadays, when a significant event first makes its appearance on social media, it will be picked up by followers who not only repost messages, but also contribute their views and opinions. An event that goes viral often has millions of people involved, leading to various discussions on issues related to it. The US election and terrorist attacks in Europe in 2016 are examples of such events.

The use of social media on such significant events offers excellent opportunities to understand the factors that may make an event go viral, the evolution patterns of events, and the reasons that motivate people to be involved in discussions. This is because social media data is multivariate and has rich information derived from the posting and reposting interactions among people. This information includes message contents, people, people's relationships, and time. However, the complexity of social media data also poses challenges to obtaining deep insights into events and their developments. The first challenge faced is the large number of people involved in significant events. When an event is retweeted hundreds of thousands of times, as seen in the 2016 US election, analyzing their tweets and retweets would require great efforts. Another challenge is related to the diversity of the contents added into the reposted messages, which may include simple comments or heated debates. The different semantics of the contents demand appropriate tools to identify and summarize themes. Furthermore, the development of an event may be unpredictable and surprising. Political events may stir more diverse discussions than sporting events, and may even lead to subsequent events. New tools are therefore needed to analyze such diverse events and their developments.

To understand people's behaviors related to significant events from social media data, we need to have a proper tool to identify and analyze the dynamic information diffusion process based on users'

2017 IEEE Conference on Visual Analytics Science and Technology (VAST)
October 1–6, Phoenix, Arizona, USA
978-1-5386-3163-8/17/$31.00 ©2017 IEEE

reposting behaviors. Efforts have been made to analyze social media events from multiple perspectives. Some research investigates an event by examining the dissemination of a single message [8,42], but they face difficulties in deriving event overview and understanding event stages. To help understand event stages, some researchers explore the dynamic evolution of keywords [4], topics [18,59] and sentiments [63] during specific events. Their work, which adopts the river-based visual metaphor, clearly identifies the evolutionary pattern of an event. Such approaches usually ignore the contributions of posters and re-posters to the changes of the semantics of events and their developments.

To better support the need for the analysis of dynamic features of social media contents related to significant events, we propose E-Map, a visual analytics method, to provide an explorable summary for multi-faceted events covered by social media (Figure 1). Under our approach, representative keywords extracted from social media messages are visualized as cities, and messages related to these keywords become towns. The distances of towns to their affiliated city are determined by the time at which messages were posted. A city and its surrounding towns form a region, which represents the theme of a set of messages. Connections between regions indicate the close relationships among the themes they represent. Rivers going through different regions reflect the reposting activities that concern multiple themes. Connected regions form a continent, while an island symbolizes isolated theme(s). The temporal evolution of the cities, towns, rivers, continents, and islands portrays the multiple facets of the development of an event. Key information such as the issues discussed, key people involved, and their posting behaviors at different stages are organically fused in a spatial and temporal environment, which can be easily navigated and interacted with.

Our research makes the following contributions:

- **A Novel Visual Metaphor for Event Summary based on Social Media Data.** The explorable map provides semantic summaries of key features of an event and its development. The map fuses the multi-faceted information and enables users to explore the event development with an intuitive visual metaphor.
- **Spatial Temporal Visual Analytics of User Behaviors in Social Media.** User behaviors are encoded as spatial objects on the map. Our visual analytics system enables analyzing how the participations of key players affect the development of an event with a map metaphor.
- **Analysis of Interesting Cases on Real-world Social Media Data.** We analyze real-world events with the proposed method. Factors that affect the evolution of significant events and the patterns of discussions are revealed by our method.

The structure of this paper is as follows. Section 2 reviews related work while Section 3 introduces the social media data, the definition of an event and its features. After the design of E-Map is presented in detail in Section 4, we explain how users can explore the multi-faceted information of social media data with the E-Map tool in Section 5. In Section 6, we demonstrate the use of our tools through two case studies and finally, we conclude the paper with its limitations and future work that can be done.

2 Related Work

Social media visual analytics draws attentions from researchers recently [14,44,58]. We review related research in the field.

2.1 Social Network Visualization and Analytics

We sum up two important categories of social networks in social media visualization, namely follower network [24] and messages diffusion network [51]. The goals of visual analytics for these social networks include social communities detection [41], key player identification [10,57,60], and information diffusion analysis [47,61]. To visualize the follower network, most of existing visualization techniques focus on network structures with node-link diagrams [24], adjacency matrixs [25], and a NodeTrix representation [26]. Recently, researchers have analyzed information diffusion and reposting relationships in social media with visual analytics [8,42,51]. Whisper [8] is one of the earliest visual analytics work to represent and analyze the spatial-temporal information diffusion process. WeiboEvents [42] provides three layouts, including a tree layout, a circular layout, and a sail layout for diffusion analysis. Most of the these techniques focus on capturing the structure of a social network. However, none of them focus on the multi-faceted information diffusion combined with reposters' network to analyze the events, which is the focus of our paper.

2.2 Event Detection with Visual Analytics

Users generate content (e.g., text, images, and multi-media) to share information, give opinions, spread news and connect with others [17,30,31]. Social media can quickly reflect and affect significant events in real world. Wanner et al. summarize four types of event detection methods, including clustering, classification, statistical methods and other algorithmic methods [56]. From a visual analytics perspective, Dou et al. [18] derive multiple topics from the event and visualize them with parallel time lines. Their extended systems support users to analyze the reasons why events break out and identify the sources and related events [43,53]. Further, researchers detect hierarchical topics in events from social media with topic modeling [19] and tree-cut algorithm [16]. TwitInfo [38] applies statistical methods to detect the peak amounts of messages as events. Based on the normal patterns in information diffusion networks, Zhao et al. propose FluxFlow [62], to identify and visualize the abnormal events in social discussions with classifiers. Besides textual information, researchers also detect events and anomalies with spatial temporal information in social media [15,36,37]. Chae et al. use seasonal-trend decomposition to detect events along the time [11,12]. Thom et al. propose ScatterBlog series [6,48,49] to support event detection for situation awareness. Though these proposed methods can detect and visualize events, providing an explorable semantic summary of key features of the event development with multi-faceted information, to our best of knowledge, has not been done before. This is the research goal in our work.

2.3 Event Evolution Visual Analytics

Event evolution visual analytics focuses on visualizing sequence trends, outliers, dynamic patterns and event relationships [20]. In social media, researchers investigate the evolution of keywords, topics, multimedia and user behaviors. WordCloud [52] is a common approach used to visualize events profiled with derived keywords. To explore the dynamic features of words, Archamault et al. [4] propose ThemeCrowds, a hierarchical word cloud, to investigate trends of keywords in political events. Hu et al. [28] propose SentenTree and use the co-occurrence patterns of keywords for sporting events analysis. Beyond keywords, researchers investigate the influence and evolution of topics in social media, which summarizes the events with semantics. Xu et al. first visualize the competition behaviors among topics in the social media with a river metaphor [60]. Furthermore, Sun et al. find that the relationships among topics include not only competition but also collaboration [46]. Researchers also use interchange features [59] and glyphs [32] on the rivers to reflect user behaviors in event propagation in social media. Besides keywords and topics, multi-media streams are also frequently derived to identify event evolution [5,58]. We can observe that intrinsically dynamic features lead many researchers to use river metaphors to visualize event evolution [46,54,59,60,63], keenly illustrating the idea of growth and change of subjects over time. However, in a river design with a time axis, other information can only be shown in one dimension, limiting the visualization of multi-faceted information.

Our focus is to analyze the relationships among active users and how they influence event evolution. Thus, we choose to construct a map composed with multi-faceted event evolution.

2.4 Map-like Visualization for Network Data

Map is a traditional, familiar and well-understood way to represent spatial-temporal data. There is a large number of work dealing with spatial temporal visual analytics [2, 3]. The focus of our paper is to use the map visual metaphor to represent the non-spatial information, especially for network data in social media. Researchers propose GMap, an interactive visualization that transforms social networks into maps to highlight communities [22] and visualizes dynamic network data [27, 39]. While these works are good at presenting mental maps in a naturalistic and humanistic approach through map-like visual metaphors, they focus less on social media data. In the field of social media, Cao et al. [9] project a social interaction graph onto a triangle map showing the multi-dimensional attributes. Similarly, Gansner et al. propose the dynamic map generation techniques for Twitter data [23]. Later, Liu et al. propose a compact rectangle-map generation method for better visual search and space saving [34]. Our previous work, D-Map [13], introduces a hexagon-based map visualization to analyze user-centric information diffusion patterns among social-media users. These visualizations are capable and intuitive. However, there is no map design targeting the analysis of significant events in social media. The complexity of event evolution analysis requires a novel structured visual representation, which is our research motivation.

3 Data Description

In this section, we introduce the data, provide a definition of an event in the context of social media, and discuss our event features.

3.1 Social Media Data

There are several popular social media platforms, including Twitter and Sina Weibo. We use Sina Weibo data, but our approach can be applied to other social media platforms. Several popular social media platforms exist, including Twitter and Sina Weibo. In our research, data from Sina Weibo is used, but our approach can be applied to other social media platforms. By manually extracting popular keywords and hashtags discussing specic events, we use these words and hashtags as query terms to generate events of interest. Social media allows users to emphasize the theme of their messages by using hashtag. A hashtag is often associated with real-world events and can help people search for relevant messages quickly. For example, when discussing the White House executive order to restrict the entry of people from some countries into the US, many social media used the #travelban hashtag. Despite its popularity, hashtag is not used in every message. For messages without a hashtag, keyword-based search is still the primary method to locate them. For example, using the query "travel ban" can give us messages that also discuss the executive order.

Our crawler gathers messages that contain the keywords and hashtags and searches for reposting messages rooted from those source messages. Each message contains information such as the timestamp, message ID, ID of the reposted message, content, and user ID. A hierarchical structure is built to reflect the reposting relationship among messages: a reposted message is a parent, while all messages reposting it are children. With the source data and the message hierarchy, we further analyze the temporal patterns, user behaviors, information diffusion patterns and semantics of an event.

3.2 Definition of Event

In our research, an event is defined as a social phenomenon or a real-world story that is associated with a hashtag or a keyword in social media. Our interest in an event also includes its development. We see an event dynamically changing with multiple stages, and involving many people who post and repost messages related to the event. In each stage, what people discuss can be about the event itself, or other topics derived from the event. The change of topics and the development of event stages are shaped by people's messages and behaviors.

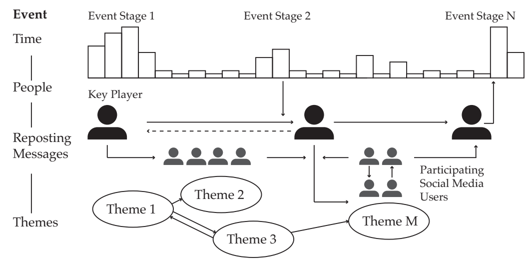


Figure 2: An illustration of an event, which consists of multiple stages. Posting and reposting behaviors lead to thematic changes.

In this sense, an event can be represented by a set of features that reflect what is discussed, who is involved, what the relationship of involved peoples is, and when the event happens. Based on Dou et al. [18], which discussed an event based on four features, $< Topic, Time, People, Location >$, we extend the event definition with four features extracted from social media data: $<$ **Time, People, Reposting Messages, Themes** $>$. These characteristics together construct a social event (Figure 2). In our discussion, "Theme" is a general term, which can be represented in different levels of semantics, including keyword and topic levels.

Both definitions above take time, people, and topics/themes into consideration. In our definition, we include messages to make full use of the richness of social media contents while dropping the location feature because the embedding location information in messages is optional on many social media platforms.

3.3 Event Features

We summarize the features of an event from four perspectives:

- **D1: Multiple Stages:** New event stages may emerge with additional new materials. Additionally, subsequent events can be generated by discussions, by new opinions, or by new participants.
- **D2: Influences from People:** Key players, such as opinion leaders with a large number of followers, can shape the discussion of an event. Besides, victims or dissidents of an event can also shift the development of the event. All these people are critical to the development of an event.
- **D3: Posting and Reposting Behaviors:** The behavior of continuous posting with specific hashtags and keywords exposes an event to the public. Reposting behaviors lead to information diffusion with new opinions and materials generated.
- **D4: Dynamic Changes of Thematic Discussions:** Discussion themes can merge, split, disappear, or resurface in different stages of an event. The themes of discussion also vary in different participating users and event stages.

4 E-Map

In this section, we discuss the motivation, design requirements for analyzing events, visual encodings and steps of map construction.

4.1 Motivation: Semantic Map

There are two types of simple presentations to visualize dynamic events, that is, the network-based representation (e.g., node-link diagram) and the space-filling based representation (e.g., Voronoi, treemap). The network-based representation shows the connections among objects but wastes blank regions. Moreover, it leads to the hair-ball clutter when there is a large number of nodes. Space-filling based representation makes full use of space and visualizes the spatial distribution of objects. However, it fails to present to connection between objects, making it hard to identify object relationships. Therefore, we need a visual design combining these advantages while providing semantic understanding.

We choose to use a map metaphor for our design for two reasons. First, a map offers a structured and semantic space for organizing information. Events and event features in our research are all based on social media messages, which are usually not structured in a way that supports the analysis of the themes and stages. A map can exhibit various types of information through different map components, including object shape, map color, and terrain. Mapping relevant information into a 2D surface with certain semantic structures can help people better understand events. Second, a map is a graphical representation that people are familiar with. People are familiar with the objects on a map, such as land, city, river, island, and continent, as well as the spatial relationship among them. Users may find it easy to learn and use map-like visualization tools if a map is used appropriately in the design.

4.2 Design Requirements

Our overall goal is to support the visual analysis of events based on social media. We will design a visualization overview to help users gain a high-level understanding of an event, as well as in-depth analytical tools that can be used to explore multi-faceted features of an event. Here, we summarize our design requirements:

- **R1: Providing a view for event summary with semantics and temporal trends**. Multiple thematic discussions should be identified and visualized (**D4**). This view can guide the in-depth analysis of different event stages. (**D1**).
- **R2: Grouping messages based on their similarity in discussing themes and reposting relationships**. The distributions of keywords extracted from messages should be visualized to help users understand the themes, their popularity (**D2**) and reposting relationship (**D3**) .
- **R3: Detecting and presenting themes in different time periods**. Keywords in different time periods should be visualized to help users see the change of themes (**D4**), and to infer the thematic evolution of an event at different stages (**D1**).
- **R4: Identifying behavior patterns of social media users in posting and reposting messages**. How key players' behavior influences event evolution should be identified (**D2**). Further supports for identifying these key players, their discussions, how and when they are involved should be provided (**D3**).

In our research, we build a semantic map to meet the above design requirements R1 to R3 and integrate various interactive tools on the map (Section 5) to support R4, as well as R1 to R3.

4.3 Semantic Map Design and Visual Encoding

Map Overview: E-Map is built with social media messages. Both the original messages and information derived from them are used for map construction (Figure 3). Event features can be mapped to map features with semantics, including cities, towns, regions, rivers, and continents (**R1**).

City: A city encodes a discussion keyword (Figure 3 - keyword city). These keywords are derived from the most popular words in social media messages in different time periods of an event. The size of a city corresponds to the number of messages containing the keyword. A city is the center of a geographic region, as the top keyword is the thematic center of messages. The distances of cities are determined by the closeness of keywords in the reposting relations and the sequential temporal relationship. Thus, users can observe the temporal trends of keywords on the map (**R3**).

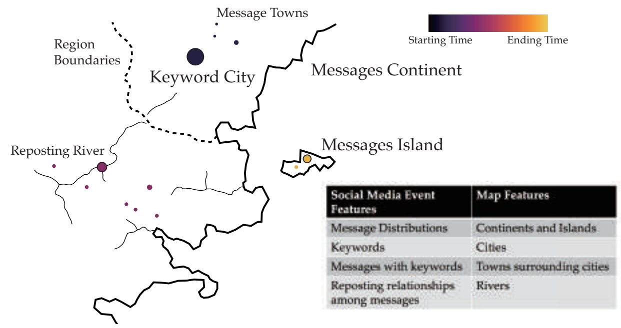


Figure 3: Visual encoding of the map. The map is constructed from social media messages. Key features of an event, including discussion themes, involved messages, and reposting behaviors are mapped to cities, towns, and rivers, respectively. Continents and islands are formed based on the similarity and connection of cities and towns. Any feature with a temporal attribute is color-coded and color mapping is based on the "inferno" perceptually-uniform color scheme [1].

Town: Towns are messages with high affiliations to a specific city (Figure 3 - message town). Towns are located around cities. The distance between a town and its city encodes time interval. The first message with the keyword is put at the center of the city, while the newest message stays farthest from the city. Thus, users can identify the message distribution regarding keywords and time (**R3**). To make the design more scalable, messages within nearer time stamps can be merged into one town, the size of which encodes the message number.

Region: Cities, their affiliated towns, and the territories between them form regions. A region has only one city (one discussion keyword), but many affiliated towns (messages). The size and shape of a region are decided by the locations of the city and towns inside it. Dashed lines specify region boundaries. The connections of multiple regions reflect the closeness of the reposting relationships and temporal relationships among the messages with the keywords that these regions represent (**R1**).

River: A river in a map symbolizes the connection of heavy reposting relations between regions. A river starts from a source region and ends in a target region, indicating that there are a large number of people who repost messages containing the keyword of the source region and whose messages also have the keyword of the target region. Thus, a river provides the hint of information diffusion and the evolution of discussions, based on features extracted from the reposting behaviors of social media users (**R4**).

Continent and Island: Large connected regions form continents, while smaller, isolated regions create islands. The connection of regions reflects information diffusion in terms of discussion keywords. The layout of continents and islands, with cities and towns, shows the distribution of themes and the multi-faceted relations (**R1**).

With regards to color mapping, we use color to encode time information because the temporal evolution of events (**R3**) is an important target of analysis. We have several design constraints. First, the color map should allow users to perceive the temporal change and sequence of relevant event features. Second, the mapping should be sequentially and perceptually uniform [50]. Lastly, the mapping applied to different map features should be consistent. Based on these requirements, we choose the "inferno" perceptually-uniform color scheme [1] (Figure 3). We apply this color coding into cities, towns, and regions (Figure 1). Alternative color schemes that meet the requirements can also be used.

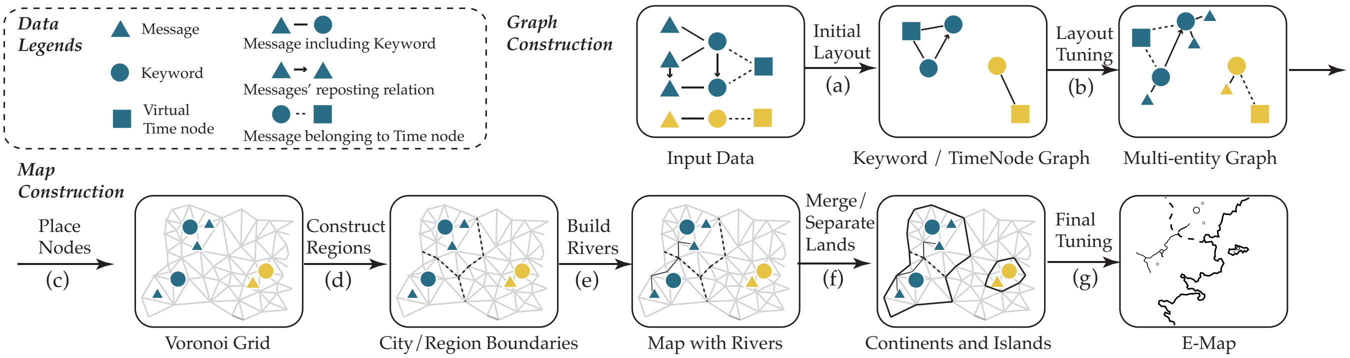


Figure 4: E-Map Construction. With the social media messages M, we obtain the keywords K and the virtual time nodes T. We layout the nodes K and T (a), and further tune the layout with M (b). With the multi-entity graph as input, we build a Voronoi grid and project nodes as cities and towns (c). Then, we calculate the regions and boundaries (d). Based on the reposting relationships, we create rivers to connect relevant regions (e). Finally, we merge the connected regions to make a continent or an island (f), and tune the map effect with erosion and smoothing processes (g).

4.4 E-Map Construction

E-Map construction includes two steps: graph construction and map generation (Figure 4).

4.4.1 Multi-Entity Graph Construction

According to our definition of the event, time and discussion themes are important features. The input is social media messages $M = (m_1, m_2, ..., m_i, ..., m_{n_m})$, while the output is the multi-entity graph $G = (N, E)$, where $N = (K, M, T)$ includes discussion keywords K, messages M, and virtual time nodes T. The edges have three types of relations $E = (KK, KT, KM)$, namely keywords to keywords KK, keywords to time nodes KT, and messages to keywords MK.

Keyword Extraction within Time Periods: The themes dynamically change as multiple people discuss different themes at different time periods. We first derive keywords from the original text in all messages. In the analysis of events, we not only care about high-frequency keywords but also have an interest in high-frequency keywords in certain time periods. Based on this motivation, we split the time into different periods and calculate the Term Frequency-Inverse Document Frequency (TF-IDF) in each period. We extract the keywords with top TF-IDF values in each period. By merging the duplicated keywords, we get a keyword set $K = (k_1, k_2, ..., k_i, ..., k_{n_k})$. In addition, we construct a virtual time node set $T = (t_1, t_2, ..., t_i, ..., t_{n_t})$, which indicates the span of time periods. We also add edges K_iT_j between the keyword K_i and their associated time node T_j.

Reposting Relationship Mapping for Keywords: For each keyword, we can obtain all messages containing the keyword. We extract the reposting relationship from all messages. If a message with a target keyword K_i reposts a message from a source keyword K_j, an edge K_iK_j is added between the pair of the keywords.

Two-round Multi-Entity Graph Layout: Based on the above operations, we get an initial graph $G' = (N', E')$, where $N' = (K, T)$ and $E' = (KK, KT)$. We initially apply the force-directed layout algorithm [21] with the simulation of stable results [29] to graph G' (Figure 4a). It makes the layout results of multiple runs of simulation similar and ensures the final map layout consistent for the same event. The goal of the initial layout is to reflect the reposting relation (KK) and to pull the keywords together within the same time periods (KT). However, there may be nodes that are too close to each other. To make node distribution more uniform, we employ the Lloyd algorithm [35] to adjust the locations of those points with a short distance in the second round of layout tuning. In addition, we add node M_i and the edge M_iK_j between the message node M_i and the keyword K_j to recalculate the layout (Figure 4b). The rule for adding the edge MK is to connect the most representative keywords of the message nodes so that the construction complexity can be reduced and the positions of message nodes are interpretable. Important keywords with a large number of messages gain larger space and push away other nodes. Furthermore, to maintain the general G' layout, we use smaller charge and strength of nodes M and edges MK compared with forces in G'. The general force can be represented as:

$$f_a = d^2/k, k = \sqrt{area/n} \tag{1}$$

where d is the distance between two nodes, k represents the ideal distance of links and n represents the corresponding nodes' number.

Thus, the positions of cities (keywords) and towns (messages) are determined in a multi-entity graph G as the input of the map layout.

4.4.2 Map Layout

The input of map construction is the multi-faceted graph with calculated positions. With the input graph G, the map layout step produces its output, an E-Map with cities, towns, regions specified by boundaries, rivers, continents, and islands (Algorithm 1). The basic data structure used in this step is a Voronoi grid.

Splatting on a Voronoi Grid: (Figure 4c) Initially, according to the screen size, we randomly project the seed points in a 2D space. More seed points lead to more precise boundaries in a map design. In our implementation, we choose 4,096 points, considering the trade-off between visual aesthetics and computational efficiency. We apply the Lloyd algorithm [35] to generate a relatively uniform distribution of points. With these points, we construct a Voronoi grid and use it as the base for map terrain. Several requirements are considered when we choose the Voronoi grid. First, the polygon in Voronoi is irregular, similar to the real-world terrain. Second, from the algorithmic perspective, the neighborhood representation is convenient to maintain based on Delaunay Triangulation. Based on the input positions of keyword nodes, message nodes and sampled points from reposting links derived from the graph G, we make a splatting of each node to the Voronoi triangles with a Gaussian Kernel (Algorithm 1 step-c). The Density (D) of each point in Voronoi with position (P) can be calculated as:

$$\sum_{n=1}^{N_k} \alpha_n \mathcal{N}(P|\mu_k, \sigma_k^2) + \sum_{m=1}^{M_k} \alpha_m \mathcal{N}(P|\mu_m, \sigma_m^2) + \sum_{l=1}^{L_k} \alpha_l \mathcal{N}(P|\mu_l, \sigma_l^2), \tag{2}$$

in which N_k, M_k and L_k are the number of keyword nodes, message nodes and sampled nodes from reposting links. α is the weight of each feature, which corresponds to the count of nodes. μ and σ are the position of the node and radius of the Gaussian Kernel of each node, respectively. The splatting radius of a node is determined by the number of messages it contains. It is also constrained with

Algorithm 1 Map Layout Algorithm

Input:

A list of keyword nodes K_i with position $K_i.pos$, and size $K_i.size$, $i = 1,2...n_k$;
A list of message nodes M_i with position $M_i.pos$, and size $M_i.size$, $i = 1,2...n_m$;
A list of keyword message edges K_iM_j; $i = 1,2...n_k$, $j = 1,2...n_m$
A set of reposting links among keyword nodes L_i with the source $L_i.source$ and the target $L_i.target$, $i = 1,2...n_l$;

Output:

A list of Voronoi grid points V_i, $i = 1,2...n_v$; A list of city points C_i, $i = 1,2...n_k$; A list of town points T_i, $i = 1,2...n_t$; A list of boundary paths for regions B_i, $i = 1,2...n_b$; A list of river paths R_i, $i = 1,2...n_r$; A list of boundary paths for continents and islands P_i, $i = 1,2...n_p$;

//Step c: Constructing the Voronoi Grid and Conducting the Splatting
grid points V = Lloyd(new Points(v)) and set all $V_{index}.density = 0$;
for $i = 0; i < n_k; i++$ **do**
  $index$ = FindMeshGrid(V, $K_i.pos$) //Find corresponding grid index in Voronoi
  $V_{index}.size = K_i.size$
  for $j = 0; j < V_{index}.neighbors.length; j++$ **do**
    $V_{index}.density$ += Gaussian($V_{index}.pos$, $V_{index}.neighbors[j].pos$, $K_i.size$)
  end for
end for
Repeat the splatting process for message nodes N_m and sampled points from L_i
//Step d: Constructing Cities, Boundaries, and Towns
Initialize city points C, boundary paths B and town points T
for $i = 0; i < n_k; i++$ **do**
  $index$ = FindMeshGrid(V, $K_i.pos$)
  $C_i.pos = V_{index}.pos$ //Constructing cities
end for
for $i = 0; i < n_v; i++$ **do**
  $cityIndex$ = FindNearestCityPointIndex(C, $V_i.pos$)
  $C_{cityIndex}.area$.push(V_i) //Set city regions
end for
for $i = 0; i < n_v; i++$ **do**
  if !isAllEqual(V_i.neighbors.cityIndex) **then**
    B.push(V_i) //Boundary node's neighbors are not the same
  end if
end for
Connect all the connected boundary nodes as final boundaries B
Build $T.cityIndex$ according to link KM
for $i = 0; i < n_k; i++$ **do**
  $C_i.timeMapping = \{domain: C_i.timeRange, range: [0, C_i.area]\}$
end for
for $i = 0; i < n_m; i++$ **do**
  $T_i.pos = C_{T_i.cityIndex}.timeMapping(T_i.timeRange)$ //Constructing towns
end for
//Step e: Constructing Rivers
Initialize river points R
for $i = 0; i < n_l; i++$ **do**
  ($source$, $target$) = FindMeshGrid(V, $L_i.source$, $L_i.target$)
  $R_i.curve$ = BuildCurves($source$, $target$) and sample grid points on the curves
end for
Sort the river points R based on visiting grid points and merge R with similar visitings
//Step f: Constructing Continents and Islands
P = Contour(grid points V, $seaLevel = 0$) //Regions inside P form lands
//Step g: Final Tunning
P = Erosion(P) //Update boundary paths with erosion effect
for $i = 0; i < n_r; i++$ **do**
  for $j = 1; j < R_i.nodes.length-1; j++$ **do**
    $R_i.nodes_j.pos$ = Interpolate($R_i.nodes_{j-1}.pos$, $R_i.nodes_{j+1}.pos$)
  end for
end for

a maximum ratio r of the width of the map. We choose $r = 1/10$ after tuning it. The value of all the triangles in the Voronoi grid is normalized for later calculation of continent and sea.

Constructing Cities, Towns, and Regions: (Figure 4d) We place the cities, as well as the keywords they represent, in the Voronoi grid, based on their positions defined in the layout graph from the previous step. The size of a city is encoded by the attaching number of messages. To achieve the region calculation, we iterate all the triangles in the Voronoi grid and assign each triangle to its nearest city. Thus, we can detect the boundaries by identifying the grid triangles with neighbors belonging to a different city (Algorithm 1 step-d). Within each region, we reorder the message towns and project them from the center city to the region boundaries based on the time they are posted. We provide the aggregation mode to summarize messages with nearer time stamps into one bigger town. The size encodes the number of messages, which is coherent to the scale of the city node. After the boundaries are calculated, we map the positions of the towns. Thus, the time encoding of the town positioning identifies a relative time. A message town comes along much later than the first occurrence of a keyword city and will be positioned on the near boundary of the city region.

Constructing Rivers: (Figure 4e) Rivers indicate the reposting relationship among different regions (keywords). With a source-target keyword pair from KK in E of the graph G, we connect all the centers of its passing triangles in the grid to construct a river. There are a lot of rivers calculated. After sorting based on reposting counts, we filter the influential repostings with a threshold. Although we filter the rivers, the crossings of rivers still lead to visual clutter. To further reduce the clutter, we merge the rivers sharing parts of the route (Algorithm 1 step-e). To decide the threshold, we mainly consider the visual complexity of the final map. We keep 5 - 10 rivers in the overview and the details of reposting behaviors can be explored dynamically on demand to achieve scalability.

Constructing the Continents and Islands: (Figure 4f) Based on the density value after the splatting of cities, towns, and rivers, we can obtain the value distribution of the triangles in the Voronoi grid. By merging those polygons with density values larger than 0, we have continents. Polygons with a density value of 0 become sea and are merged as well. Thus, some isolated continents are regarded as islands. Outlining the contours of sea level in the Voronoi grid, we obtain the boundaries of continents and islands (Algorithm 1 step-f).

Final Tuning of the Map: (Figure 4g) Till now, we place all the features on the Voronoi grid. The final tuning includes the following steps. First, we simulate the erosion process with the Planchon-Darboux algorithm [40]. The process is to change the shape of landforms by simulating the fluvial erosion. Thus, we can get a real-world like map shape along the continent boundaries. Second, the original rivers and boundaries tend to be in a zig-zag fashion along the Voronoi grid. To make them smoother, we relax the positions of those middle points in the paths and interpolate the relaxed positions of their upstream and downstream neighbors. This step results in more natural-looking rivers and boundaries (Algorithm 1 step-g), and eventually, an E-Map that looks like a real map.

The complexity of most algorithmic processes is $O(n)$, except that the algorithm for tree traversing and sorting processes in the preprocessing stage is $O(n \log n)$.

5 Spatial Temporal Visual Exploration

We propose a visual analytics pipeline for the exploration of the evolution of an event (Figure 5). It is achieved by visual analytics system to support the exploration of the event map (Figure 6).

5.1 Spatial Temporal Overview

We provide a spatial temporal overview to help users quickly get the big picture of an event. It shows the distribution of important keywords related to the events (Figure 1a). Cities with larger size indicate the widespread participation. The neighboring cities have tight connections, in both their reposting relationship and reposting time. The distribution of the continents and islands also indicates the connections among different themes. We provide several buttons enabling users to toggle the visibility of map features, including density representation, cities, towns, rivers, and region boundaries (Figure 6a).

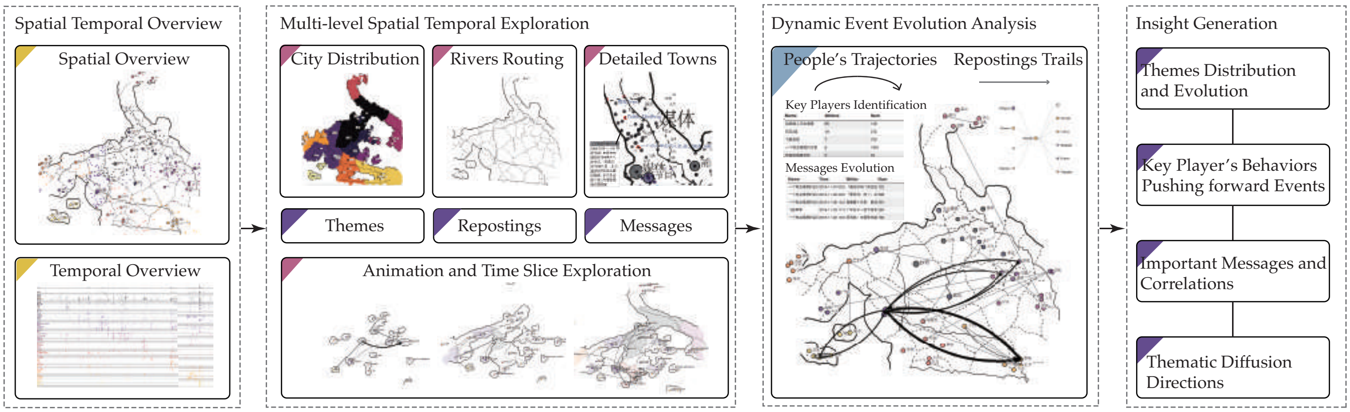


Figure 5: Visual analytics pipeline of E-Map. Initially, a spatial and temporal overview is provided. Users can conduct multi-level spatial temporal exploration, including navigating, lasso selection, and time filtering, to examine the details of the map. Time slicing and animation are supported to help users understand how these maps are dynamically constructed, and how an event evolves. To further understand the development of the event, users can investigate the behaviors of people using social media, such as continuously posting and reposting. These behaviors are encoded as people's trajectories and reposting trails. Based on these explorations, users can gain insight into an event and its evolution.

The temporal view (Figure 6b) allows users to examine the temporal patterns of individual keywords. Keywords of an event are lined up vertically and ordered by the time they first appear. The keywords are color-coded by their peaking time of appearance to aid in their comparison. Figure 6b shows an order with the earliest keywords on the top, but the order can be reversed interactively. The X axis is the timeline. Each keyword takes a row, and the blocks on each row indicate the appearance of keywords in messages. Both the height and color intensity of the block are used to encode the number of messages containing the keyword. With this overview, users can get the hints of keywords bursting at a different time and gain insight into the evolution patterns of the event.

5.2 Multi-level Spatial Temporal Exploration

City and town are two levels of spatial representations, forming the hierarchical structure as seen in real maps. To support multi-level spatial temporal exploration, we provide three types of interactions.

First, users can brush the temporal view. The selected messages with the specific time range reconstruct the E-Map. The new map is a subset of the original map and highlights the keywords and messages in the selected time range. Thus, it helps users better examine the detailed event features in the selected time ranges. Further, users can animate the map construction process between two specified time points (Figure 6g). The expansion of continents is one important feature of E-Map. With the expansion metaphor, users can perceive the event evolution with the variance and emergence of themes. To help users preserve their mental map, we provide a smooth transition for animating the thematic development processes of regions, continents, and islands.

Second, users can explore the map in a familiar way as they do with online real-world maps. We provide navigation, zooming, and clicking selection for in-depth analysis. When users drill down, the detailed towns are automatically shown to reflect the hierarchical structure of the spatial organization. All features including cities, towns, and rivers are clickable. To further support multi-level exploration, we provide two linked information panels, a message panel (Figure 6b) and a key player panel (Figure 6c). Users can select a keyword city to highlight all related messages. They can also select a particular town to see the raw message content, and when the message is posted and by whom (Figure 6e-S1, S2, S3). Linked with the map, the message panel provides the related messages with timestamps, texts, reposting numbers and the participants. The key player panel shows the statistic information, including how many messages an individual participant has posted and reposted.

Third, users can apply a free polygon brush on the map, to select groups of towns for analysis. Each region only uses the most popular keyword to label its center city, and other less popular keywords are invisible. After brushing regions, a word cloud will appear on the map to show these keywords within their region (Figure 1c). The size of a keyword is determined by its frequency of occurrence.

5.3 Event Evolution Analytics

To further reason why the map changes, we propose event evolution analysis based on human behaviors. We aim to identify the event stages with summarized keywords, the behaviors of key players and their important messages. There are two important features in exploring event evolution on the map: the trajectories of social media users and their connections. On one hand, the person who posts multiple messages would move around different towns on the map. Thus, trajectories on the map reflect the posting and reposting of messages with different themes. By summarizing the aggregated movement patterns on the map, we can tell the stories of the main trends in the thematic changes of discussions. On the other hand, messages posted by different persons can be reposted by others. Each town can be connected with other towns because of such reposting connections. These connection behaviors with high influence have been summarized as rivers. In the E-Map, we use black curve routes to encode the trajectories and use straight gray links to encode the reposting connections. The thickness of the links encodes the number of messages (Figure 6a).

We provide two modes for visualizing the trajectories and reposting connections: an individual mode and an aggregated mode. In the individual mode, links, including both trajectories and connections, are shown at the town level to represent the relationship among messages. In the aggregated mode, links are summarized at the city level by aggregating all messages of towns affiliated with the particular city (Figure 6a). These two types of links are important because they help to identify how the event changes and what people may influence. By sorting the reposting number, users can explore the messages with high impact. In addition to the aggregation, we add the filters and toggles for the trajectories and connections, respectively, to further reduce clutter. By brushing the temporal view, users can see the appearance, disappearance, or assimilation of regions with these trajectories, which offer the evidence of event evolution. To enable detailed exploration, we provide a one-dimensional sequence view to visualize the trajectory of important users (Figure 6f). Each rectangle represents a city that an individual user visited. The x-axis reflects the temporal order, and the height represents the message amount from a specific visit. The view is linked with others and reflects the temporal trends of user behaviors.

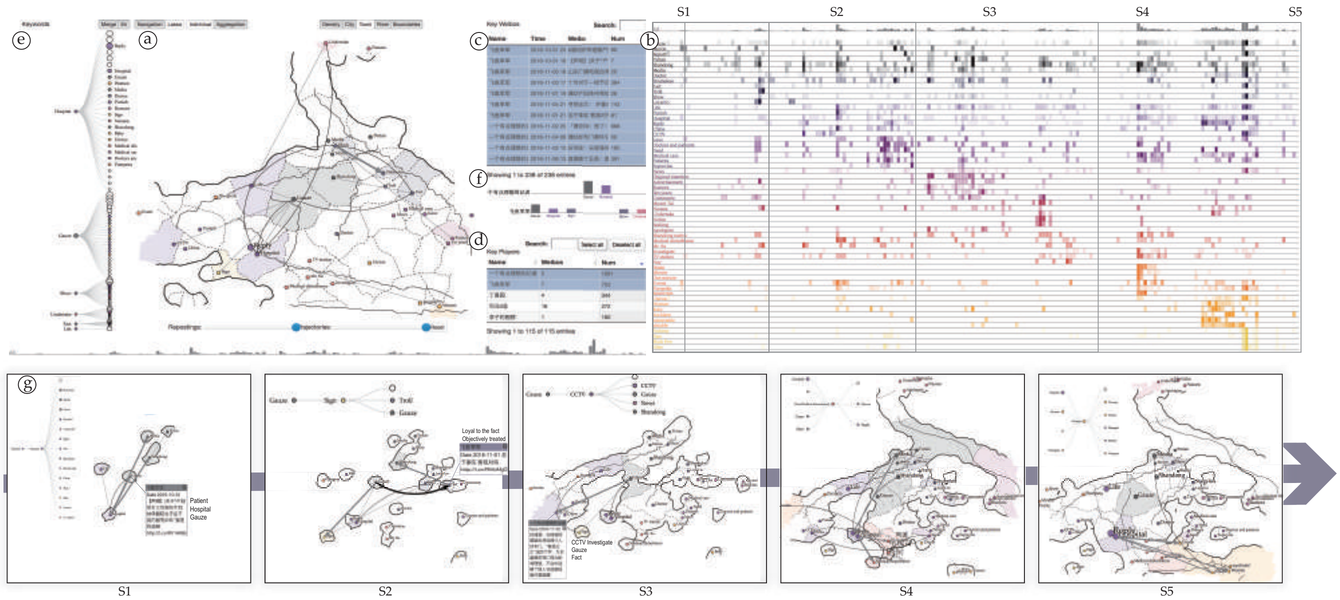


Figure 6: Spatial temporal visual analytics system of E-Map, including (a) a map view, (b) a temporal view, (c) a message panel, (d) a key player panel, (e) a detailed keyword relationship view. and (f) a sequence view. (g) Time slicing and animation are provided for analyzing event evolution. In this case, five event stages are identified and each stage has its key players. We can analyze how they affect and shape the event development.

To identify the processes of theme development, we provide a detailed keyword reposting view (Figure 6e). Users can brush a region on the map, or select cities or the trajectories to feed information to the keyword reposting view. There are three columns: the keywords from the selected messages are in the middle, and their parent and child keywords in the reposting hierarchy are on the left and right, respectively. The size of a circle is tied to the number of messages containing the keyword. This view can show the information diffusion process involving different keywords (Figure 6e).

In summary, E-Map enables users to explore multi-level details of the dynamics of an event and its evolution at a different level of granularity. The implementation is based on D3.js [7]. We also developed an internal dictionary to support both Chinese and English.

6 Case Study

We present two cases from real-world events in Sina Weibo and demonstrate how our system can help to analyze event evolution.

6.1 Case 1 - Gauze Scandal: A Social Case

In this case, we illustrate an exploration process using the E-Map to understand a bursting event by identifying the event stages, key players, reposting relations and themes evolution. This event is a real-world story starting from a Weibo message, posted by a TV Host, claiming that a woman in August 2016 was left with a piece of gauze inside her body by doctors after surgery. The story was later picked up by the new media, which led to heated discussions on Sina Weibo in October-November 2016, many of which criticized the hospital and doctor involved. However, as more people joined and the discussion became broader, new information was revealed to tell an entirely different story (Figure 6). We used the query "Gauze Scandal" (in Chinese) to crawl the Sina Weibo messages. Using the retrieved messages as the seeds, we crawled all messages reposting them. In total, we got 6,963 people and 9,626 messages. Figure 6 shows the procedure of using our E-Map for this analysis. By default, we select the top ten keywords of each time range. After merging keywords and filtering out non-text messages, we finally obtain 60 keywords and their spatial temporal distributions (Figure 6b). We can see that the keywords such as "*Gauze, Life, Shandong, Media, Hospital, Doctor, Broadcast*" are initially highlighted and continuously discussed (Figure 6a). These keywords outline the beginning of the event. To further analyze the event, we split the timeline and analyzed the event step by step (Figure 6g S1-S5).

At the first stage, S1, from October 30th to 31st, "Junjun", the host of a local TV show, "Life Help", in Shandong Province, which first picked the story, posted several messages about *"Gauze"*, asking questions such as "When can we get the gauze out?" (Figure 6g-S1). These were reposted by only a few people. The regions appearing at this stage are largely about the story, the media outlet, the show, hospital, and place. At the next stage S2, November 1st to 2nd, "Junjun" continuously pursued this story and emphasized it was a *"Fact"*, as indicated by a new region of *"Fact"* and a curved link between *"Gauze"* to *"Fact"*) (Figure 6g-S2). However, other people such as "A journalist with dream" stood out and argued that the hospital and doctor were real victims and the patient with gauze inside was rude and unreasonable. He provided evidence about *"Signing"*, arguing that "it was because the woman didn't want to sign the surgery to take the gauze out". He led the discussion theme to *"Signing"*, as indicated by the gray straight link of repostings (Figure 6g-S2). Further, more people joined the camp of the "Journalist" and called "Junjun" *"Shameless"*, which formed a new region. However, there were still some active supporters of "Junjun", who blamed the hospital.

At the next stage S3, November 2nd to 3rd, the discussion was reversed entirely. CCTV, the most influential TV outlet in China, investigated the story and found it was the fault of the patient, not the hospital. A new region of *"CCTV"* emerged on the west side of the map, as did a reposting link from *"Gauze"* to *"CCTV"* (Figure 6g-S3), as more people started to talk about the investigation of *"CCTV"* and suggested to *"Punish"* the patient. At this stage, several "We Media" of healthcare, such as "DingXiang Yuan" and "Journalist" made great efforts to push the facts to the public. At the next stage S4, November 3rd to 6th, discussions continued, and the public blamed "Junjun" and the local TV show for creating a *"rumor"* and for their ties with other private hospitals *"Putian Hospital"*, which had a bad reputation in China (Figure 6g-S4). At the final stage S5, November 6th to 8th, the development of the event slowed down, as shown by the relatively stable map, and discussion focused on the reflection of issues, such as *"Woman"*, *"Uterus"*, etc. (Figure 6g-S5).

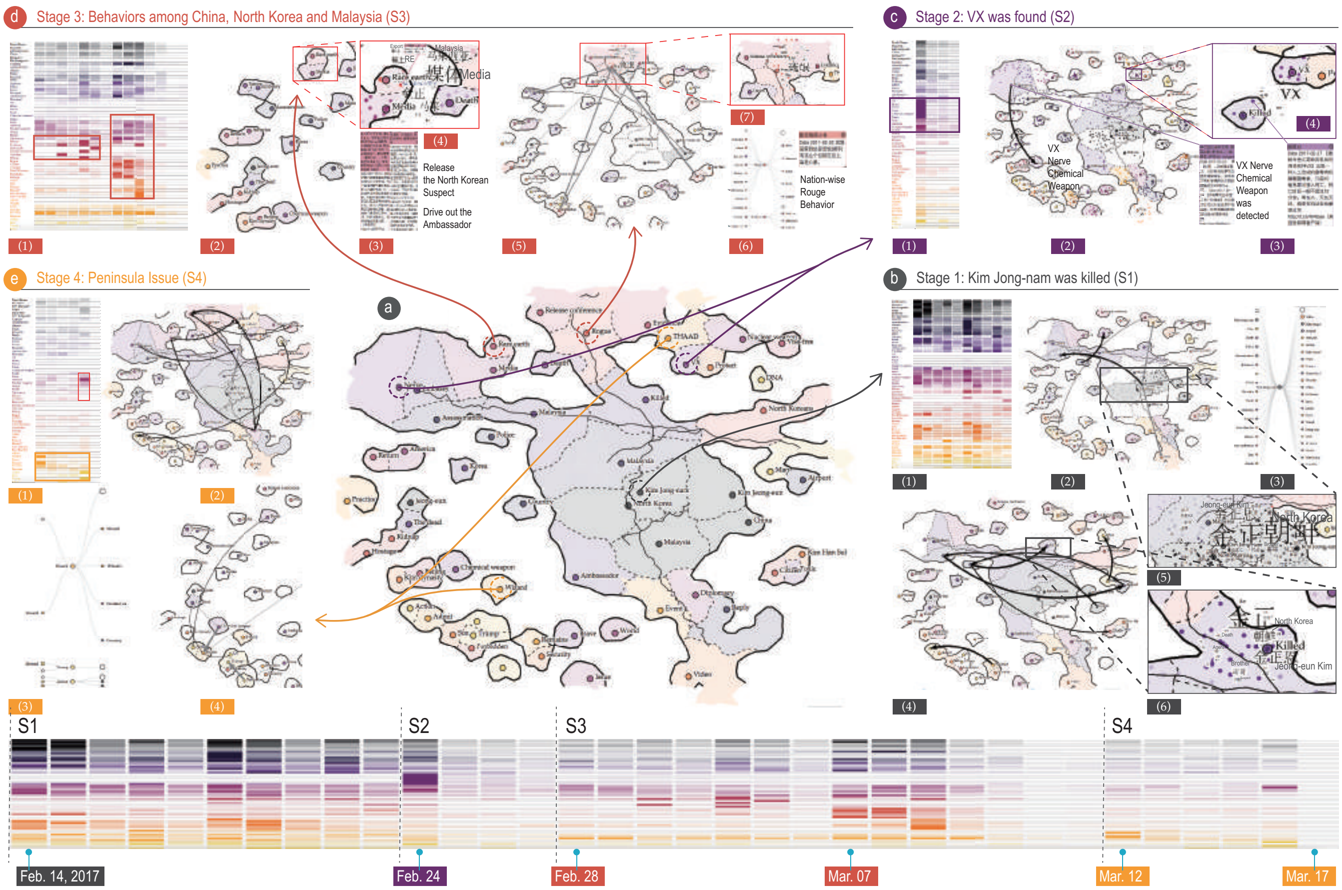


Figure 7: Multi-level spatial temporal exploration of a political event related to the death of Kim Jong-nam. E-map gives an overview of the event (a) and then four event stages are presented (b-e), with newly emerging keywords, such as *"VX, Nerve, Return, Rare Earth"*, and posting behaviors. Drilling down to the detail, the system helps to analyze how information diffuses and how public opinions change.

In this case, by analyzing individual event stages, we examine the Gauze Event and find the key players who affected the course of discussion and the details of discussions at each stage. With the metaphor of map expansion, users can easily understand the event evolution process. They can pay attention to the new and changing regions. The metaphor also indicates that more people get involved and discussions become more diverse with time. The expansion of continents, the changes of the trajectories of key players, and the appearance of reposting links all help to understand the event development and the important factors behind. Compared with it, the traditional graph or space-filling techniques can not intuitively represent such semantic features with multiple event stages.

6.2 Case 2 - Death of Kim Jong-nam: A Political Case

We analyze the event associated with the death of Kim Jong-nam. On Feb.14, a person, who died in Malaysia, was recognized as Kim Jong-nam, the brother of the current leader of North Korea. His death quickly became a significant event in social media, particularly after more evidence surfaced to show he died after a poison attack. We used the term *"Kim Jeong-nam, North Korea, Malaysia"* to search Weibo messages, and got 222,678 messages from 130,197 people. Messages were posted on February 14th to March 17th, 2017.

The map gives us an overview of the event (Figure 7). In the center, we found the keywords *"Malaysia, North Korea, Kim Jeong-nam, Kim Jeong-eun, Ambassador, Assassination"*, which are the main themes of the event (Figure 7a). Initially at Stage 1 (February 14th to 24th) (Figure 7-b1), although there was no official announcement about who was killed, nearly all news agency accounts and individuals thought the so-called *"Kim Chol"* was actually *"Kim Jeong-nam"*, and multiple accounts reposted the announcement of the incident discussed with the keywords (Figure 7-b2, b3). One type of voice was from messages like "Kim was killed by his brother, to avoid the regime snatch" (Figure 7-b5). People discussed how such *"Assassination"* happened and why it occurred in a crowded place in the *"Airport"* of *"Malaysia"* (Figure 7-b4). At the same time, the police officers of Malaysia continuously released the findings in the following days. The trajectories from *"Police"* to *"Airport"* can be identified (Figure 7-b5, b6). At the end of this stage, a new story about how Kim Jong-nam was killed was announced on February 24th (Figure 7c). There were newly emerging keywords, namely *"VX, Nerve, Toxic, Chemical Weapon"* (Figure 7-c1, c2). After the description of VX by experts, the public was astonished to see the usage of such an illegal chemical weapon (Figure 7-c3, c4).

In addition to the investigation of the assassination itself, our map reveals the complexity of international affairs with several following events at S3 and S4. During February 28th and March 7th, there were three important issues being discussed (Figure 7d). Malaysia released a North Korean suspect because they lacked the evidence to arrest him on March 4th (Figure 7-d1, d3). We could see the newly emerged keywords *"Return"*. With this keyword, we found that Malaysia drove the ambassador of North Korea out of the nation on March 7th (Figure 7-d3). Another new keyword was *"Rare Earth"* (Figure 7-d1, d2). By drilling down to the detail, we found that North Korea decided to stop the export of rare earth to China (Figure 7d4). However, the response from most Chinese was like "It's just like stopping exporting oil to Saudi Arabia". Following, the North Korea government committed another *"Rogue"* action, triggering another round of reposting, playing the blame-game (Figure 7-d6, d7). To offend Malaysia, North Korea forbade all the Malaysians in North Korea from leaving the country. The messages were picked from the keywords regions of *"Ambassador, Forbidden, Citizen"* (Figure 7-d5). Afterward, in Stage S4, more interference came from the US and South Korea (Figure 7-e1). The main issue discussed was about *"THAAD"* (Figure 7-e1), which was also related with the *"Nuclear"* issues of North Korea (Figure 7-e3). Malaysia even wanted to get a *"Wizard"* to deal with it, which was made fun of by the public (Figure 7-e3, e4). All the behaviors reflect the conflicting situation in the Korea Peninsula.

For the recent political event, we are amongst the first to construct the map and analyze the event evolution. The event is still on-going, and our E-Map can load more data to support further analysis. We can analyze the evolution of chained events and identify multiple aspects of events.

We also collected feedback from a media expert on our system. We presented our system to him and he indicated that *"It's easy to understand that the dynamic changing process of the E-Map is the process of event evolution. Such a metaphor is intuitive. Especially the interactions on such a map can help us to explore the significant events and users' behavior in social media as a whole"*.

7 Discussion

We discuss the advantages and limitations of our E-Map. E-Map constructs the semantic features in spatial temporal space to help the understanding of complex events. We address both the dynamic patterns and thematic diffusion features of the events. In particular, the support for exploring spatial temporal features in an intuitive map metaphor allows users to find key players, their behaviors (trajectories), their influences (reposting links), and event development.

	CompactMap [33]	D-Map [13]	E-Map
Theme	YES	NO	YES
Time	YES	Implicit	Implicit
Reposting	NO	YES	YES
Key Player	NO	YES	YES
Multi-level Exploration	NO	NO	YES
Event Evolution	YES	NO	YES
Visual Form	Rectangle	Hexagon	Natural

Table 1: Comparison of three map-like visualizations for social media.

We compare our E-Map with two other designs that also use a map metaphor, namely CompactMap [33] and D-Map [13] (Table 1). CompactMap explicitly visualizes the dynamic message flow with time information but lacks the support for analyzing reposting behaviors, which are important to understanding how they affect the dynamics of events. D-Map focuses on ego-centric information diffusion and considers reposting behaviors and key player behaviors. However, event evolution patterns are not addressed in D-Map. In comparison, our method integrates reposting behaviors and event evolution tightly and allows deeper analysis of the dynamics of event developments and possible causes. Also, our design offers two more advantages. Firstly, our map visual metaphor is more tightly and naturally embedded with the semantics of event features, potentially simplifying user interactions. Secondly, we leverage multi-level spatial temporal analysis methods to support the understanding of complex reposting behaviors in the context of event analysis. This approach improves the scalability of data handling.

There are some design trade-offs and scalability issues of E-Map. Firstly, the maximum number of cities and towns could be 10^4 when the map displays on a standard screen with 1920x1080 resolution. However, due to the perception scalability, the current map shows at most 10^3 keywords as cities. A filter is applied to keywords of social media messages so that the map would not be cluttered. The map cannot be infinitely large, and too many keywords can make it overcrowded. We have to balance between sufficient keywords for semantic significance and a clear map view for effective interaction and navigation. Secondly, we only assign one keyword to one message so that the location of a message is unique on the map. In reality, a message contains one or more keywords or no keyword at all. If we had taken whatever keywords a message may have in deciding its location, a stable map would be impossible to get because of the unpredictability of the location of a message with multiple keywords. Thus, we choose the keyword with the highest percentage of occurrence in a message and then project the message uniquely on the map. To compensate for the loss of information, we allow users to explore all the keywords distributions of messages with the word cloud by brushing specific regions. Thirdly, the capability of the map can currently handle 10^5 social media messages as tested in the second case. Our E-Map has a limitation in its pre-processing time. If the message amount is large, the text pre-processing time, including time to extract keywords, calculating TF/IDF and force-directed layout iteration, would limit the efficiency. After preprocessing once, users can interact with the map. We will apply parallel text mining algorithms in the future to provide on-the-fly processing. Moreover, we envision to improve the E-Map by combining the topic modeling techniques with keyword analysis. With suitable topic modeling techniques for short messages, we can replace the keywords with summarized topics to enhance the theme evolution analysis with new E-Maps.

The current E-Map design takes a heuristic approach, in which the positions of semantic features are generated based on the force-directed layout with the topological relationship of extracted messages, keywords, and virtual time nodes. However, though it can disclose the semantic similarity efficiently, it is not yet a metric space as the distance between nodes does not follow the distance requirement in a metric space strictly. The graph drawing approach would cause visual ambiguities to impede the understanding of the network structure [45, 55] (e.g. the neighboring cities might not be close in semantics). Therefore, we add the boundaries, continents, and islands as semantic features to reduce such bias. Secondly, first-time users might mistake our data as geo-tagged social media data. However, our focus is the reposting messages without geographic information. Thirdly, since our map is a visual metaphor, it is not entirely equal to the map in spatial analysis. Currently, only the distance but not the direction or orientation plays a relevant role. We envision to add restrictions on the initial layout, to pull different direction forces with semantic meaning (e.g. time or other high-dimensional features). Thus, we can further improve on the expressiveness of E-Map.

8 Conclusion

We propose a novel visualization method, E-Map, to provide an explorable map to support the analysis of an event from different perspectives. With social media messages about an event of interest, we transform messages, keywords of messages, reposting behaviors, and related keywords into such map features as towns, cities, rivers, and connected regions in a continent or an island. With the vivid map construction, we also provide spatial temporal visual analytics techniques for the exploration of the event evolution based on related features. We also presented two case studies of using our system to understand real-world events. The system helps to identify interesting patterns of event development, key players, and the ways they shape and affect the development of an event.

Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable suggestions and comments. The authors also thank Martin O'Leary and Amit Patel for the inspiration. This work is funded by NSFC No. 61672055, NSFC Key Project No. 61232012, and the National Program on Key Basic Research Project (973 Program) No.2015CB352503. This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center.

REFERENCES

[1] Inferno colormap designed by van der walt and smith for matplotlib, [Online; accessed 2017-03-31]. http://bids.github.io/colormap/.

[2] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 51–58, 2008.

[3] G. Andrienko, N. Andrienko, J. Dykes, S. I. Fabrikant, and M. Wachowicz. Geovisualization of dynamics, movement and change:key issues and developing approaches in visualization research. *Information Visualization*, 7:173–180, 2008.

[4] D. Archambault, D. Greene, P. Cunningham, and N. Hurley. Themecrowds: Multiresolution summaries of twitter usage. In *Proc. of the 3rd International Workshop on Search and Mining User-generated Contents*, SMUC '11, pp. 77–84. ACM, 2011. doi: 10.1145/2065023.2065041

[5] J. Bian, Y. Yang, H. Zhang, and T. S. Chua. Multimedia summarization for social events in microblog stream. *IEEE Transactions on Multimedia*, 17(2):216–228, Feb 2015. doi: 10.1109/TMM.2014.2384912

[6] H. Bosch, D. Thom, F. Heimerl, E. Pttmann, S. Koch, R. Krger, M. Wrner, and T. Ertl. Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics*, 19:2022–2031, 2013. doi: 10.1109/TVCG.2013.186

[7] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[8] N. Cao, Y. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE transactions on visualization and computer graphics*, 18(12):2649–2658, 2012. doi: 10.1109/TVCG.2012.291

[9] N. Cao, Y. R. Lin, F. Du, and D. Wang. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications*, PP(99):1–1, 2016. doi: 10.1109/MCG.2015.73

[10] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*, 2010.

[11] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 143–152, 2012. doi: 10.1109/VAST.2012.6400557

[12] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, and D. S. Ebert. Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics*, 38:51 – 60, 2014. doi: 10.1016/j.cag.2013.10.008

[13] S. Chen, S. Chen, Z. Wang, J. Liang, X. Yuan, N. Cao, and Y. Wu. D-map: Visual analysis of ego-centric information diffusion patterns in social media. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 41–50, 2016.

[14] S. Chen, L. Lin, and X. Yuan. Social Media Visual Analytics. *Computer Graphics Forum*, 36(3):563–587, 2017. doi: 10.1111/cgf.13211

[15] S. Chen, X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. L. Zhang, and J. Zhang. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):270–279, Jan 2016. doi: 10.1109/TVCG.2015.2467619

[16] W. Cui, S. Liu, Z. Wu, and H. Wei. How hierarchical topics evolve in large text corpora. *IEEE transactions on visualization and computer graphics*, 20(12):2281–2290, 2014. doi: 10.1109/TVCG.2014.2346433

[17] M. Dörk, D. M. Gruen, C. Williamson, and M. S. T. Carpendale. A visual backchannel for large-scale events. *IEEE transactions on visualization and computer graphics*, 16(6):1129–1138, 2010. doi: 10.1109/TVCG.2010.129

[18] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou. Leadline: Interactive visual analysis of text data through event identification and exploration. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 93–102, 2012. doi: 10.1109/VAST.2012.6400485

[19] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19:2002–2011, 2013. doi: 10.1109/TVCG.2013.162

[20] F. Du, B. Shneiderman, C. Plaisant, S. Malik, and A. Perer. Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1636–1649, June 2017. doi: 10.1109/TVCG.2016.2539960

[21] T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[22] E. Gansner, Y. Hu, and S. Kobourov. Gmap: Visualizing graphs and clusters as maps. In *Visualization Symposium (PacificVis), IEEE PacificVis*, pp. 201 –208, 2010.

[23] E. R. Gansner, Y. Hu, and S. C. North. Interactive visualization of streaming text data with dynamic maps. *J. Graph Algorithms Appl.*, 17(4):515–540, 2013.

[24] J. Heer and D. Boyd. Vizster: Visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 32–39. IEEE, 2005.

[25] N. Henry and J.-D. Fekete. Matlink: Enhanced matrix visualization for analyzing social networks. In *Proc. International Conference on Human-Computer Interaction*, vol. 4663, pp. 288–302, 2007.

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

[27] M. Hu, S. Liu, F. Wei, Y. Wu, J. Stasko, and K.-L. Ma. Breaking news on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 2751–2754. ACM, 2012. doi: 10.1145/2207676.2208672

[28] M. Hu, K. Wongsuphasawat, and J. Stasko. Visualizing social media content with sententree. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2017. doi: 10.1109/TVCG.2016.2598590

[29] T. Jakobsen. Advanced character physics. In *Game Developers Conference*, 2001.

[30] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pp. 56–65. ACM, 2007. doi: 10.1145/1348549.1348556

[31] M. Liu, S. Liu, X. Zhu, Q. Liao, F. Wei, and S. Pan. An uncertainty-aware approach for exploratory microblog retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 22:250–259, 2016. doi: 10.1109/TVCG.2015.2467554

[32] Q. Liu, Y. Hu, L. Shi, X. Mu, Y. Zhang, and J. Tang. Egonetcloud: Event-based egocentric dynamic network visualization. In *Proc. of IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 65–72, Oct 2015. doi: 10.1109/VAST.2015.7347632

[33] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE InfoVis 2013*, 19(12):2436–2445, 2013.

[34] 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*, pp. 48–55, 2013.

[35] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, Sept. 1982. doi: 10.1109/TIT.1982.1056489

[36] A. M. MacEachren, A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 181–190, 2011. doi: 10.1109/VAST.2011.6102456

[37] A. M. MacEachren, A. C. Robinson, A. Jaiswal, S. Pezanowski, A. Savelyev, J. Blanford, and P. Mitra. Geo-twitter analytics: Applications in crisis management. *Proceedings, 25th International Cartographic Conference*, pp. 3–8, 2011.

[38] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: Aggregating and visualizing microblogs for

event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 227–236. ACM, 2011. doi: 10.1145/1978942.1978975

[39] D. Mashima, S. Kobourov, and Y. Hu. Visualizing dynamic data with maps. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1424–1437, 2012.

[40] O. Planchon and F. Darboux. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. *Catena*, 46(2):159–176, 2002.

[41] K. Reda, C. Tantipathananandh, A. Johnson, J. Leigh, and T. Berger-Wolf. Visualizing the evolution of community structures in dynamic social networks. In *Proc. of EuroVis*, pp. 1061–1070. Chichester, UK, 2011. doi: 10.1111/j.1467-8659.2011.01955.x

[42] D. Ren, X. Zhang, Z. Wang, J. Li, and X. Yuan. Weiboevents: A crowd sourcing weibo visual analytic system. In *Visualization Symposium (PacificVis), IEEE PacificVis Notes*, pp. 330–334, 2014.

[43] W. Ribarsky, D. X. Wang, and W. Dou. Social media analytics for competitive advantage. *Computers & Graphics*, 38:328–331, 2014.

[44] T. Schreck and D. A. Keim. Visual analysis of social media data. *IEEE Computer*, 46(5):68–75, 2013. doi: 10.1109/MC.2012.430

[45] C. Schulz, A. Nocaj, J. Goertler, O. Deussen, U. Brandes, and D. Weiskopf. Probabilistic graph layout for uncertain network visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):531–540, Jan 2017. doi: 10.1109/TVCG.2016.2598919

[46] G. Sun, Y. Wu, S. Liu, T.-Q. Peng, J. Zhu, and R. Liang. Evoriver: Visual analysis of topic coopetition on social media. *IEEE transactions on visualization and computer graphics*, 20:1753–1762, 2014. doi: 10.1109/TVCG.2014.2346919

[47] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pp. 807–816. ACM, 2009. doi: 10.1145/1557019.1557108

[48] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Visualization Symposium (PacificVis), IEEE PacificVis*, pp. 41–48, 2012.

[49] D. Thom, R. Kruger, T. Ertl, U. Bechstedt, A. Platz, J. Zisgen, and B. Volland. Can twitter really save your life a case study of visual social media analytics for situation awareness. In *Visualization Symposium (PacificVis), IEEE PacificVis*, pp. 183–190, 2015. doi: 10.1109/PACIFICVIS.2015.7156376

[50] K. M. Thyng, C. A. G. R. D. Hetland, H. M. Zimmerle, and S. F. DiMarco. True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, 29:9–13, 2016.

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

[52] F. B. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15:1137–1144, 2009. doi: 10.1109/TVCG.2009.171

[53] X. Wang, W. Dou, Z. Ma, J. Villalobos, Y. Chen, T. Kraft, and W. Ribarsky. I-si: Scalable architecture for analyzing latent topical-level information from social media data. *Computer Graphics Forum*, 31:1275–1284, 2012. doi: 10.1111/j.1467-8659.2012.03120.x

[54] X. Wang, S. Liu, Y. Chen, T.-Q. Peng, J. Su, J. Yang, and B. Guo. How ideas flow across multiple social groups. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 51–60, 2016.

[55] Y. Wang, Q. Shen, D. Archambault, Z. Zhou, M. Zhu, S. Yang, and H. Qu. Ambiguityvis: Visualization of ambiguity in graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):359–368, Jan 2016. doi: 10.1109/TVCG.2015.2467691

[56] F. Wanner, A. Stoffel, D. Jckle, B. C. Kwon, A. Weiler, and D. A. Keim. State-of-the-art report of visual analysis for event detection in text data streams. In R. Borgo, R. Maciejewski, and I. Viola, eds., *EuroVis - STARs*. The Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141176

[57] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, pp. 705–714, 2011. doi: 10.1145/1963405.1963504

[58] Y. Wu, N. Cao, D. Gotz, Y.-P. Tan, and D. A. Keim. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia*, 18:2135–2148, 2016. doi: 10.1109/TMM.2016.2614220

[59] Y. Wu, S. Liu, K. Yan, M. Liu, and F. Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE transactions on visualization and computer graphics*, 20(12):1763–1772, 2014. doi: 10.1109/TVCG.2014.2346920

[60] P. Xu, Y. Wu, E. Wei, T. Peng, S. Liu, J. J. H. Zhu, and H. Qu. Visual analysis of topic competition on social media. *IEEE transactions on visualization and computer graphics*, 19(12):2012–2021, 2013. doi: 10.1109/TVCG.2013.221

[61] J. Yang and S. Counts. Predicting the speed, scale, and range of information diffusion in twitter. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*, 2010.

[62] J. Zhao, N. Cao, Z. Wen, Y. Song, Y. Lin, and C. Collins. Fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE transactions on visualization and computer graphics*, 20(12):1773–1782, 2014. doi: 10.1109/TVCG.2014.2346922

[63] J. Zhao, L. Gou, F. Wang, and M. Zhou. Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *Proc. of IEEE Visual Analytics Science and Technology (VAST)*, pp. 203–212, 2014. doi: 10.1109/VAST.2014.7042496