D-Map+: Interactive Visual Analysis and Exploration of Ego-centric and Event-centric Information Diffusion Patterns in Social Media

SIMING CHEN, SHUAI CHEN, and ZHENHUANG WANG, Key Laboratory of Machine Perception (Ministry of Education) and School of EECS, Peking University, China
JIE LIANG, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
YADONG WU, Southwest University of Science and Technology, China
XIAORU YUAN, Key Laboratory of Machine Perception (Ministry of Education) and School of EECS, Peking University, China

11

Information diffusion analysis is important in social media. In this work, we present a coherent ego-centric and event-centric model to investigate diffusion patterns and user behaviors. Applying the model, we propose Diffusion Map+ (D-Maps+), a novel visualization method to support exploration and analysis of user behaviors and diffusion patterns through a map metaphor. For ego-centric analysis, users who participated in reposting (i.e., resending a message initially posted by others) one central user's posts (i.e., a series of original tweets) are collected. Event-centric analysis focuses on multiple central users discussing a specific event, with all the people participating and reposting messages about it. Social media users are mapped to a hexagonal grid based on their behavior similarities and in the chronological order of repostings. With the additional interactions and linkings, D-Map+ is capable of providing visual profiling of influential users, describing their social behaviors and analyzing the evolution of significant events in social media. A comprehensive visual analysis system is developed to support interactive exploration with D-Map+. We evaluate our work with real-world social media data and find interesting patterns among users and events. We also perform evaluations including user studies and expert feedback to certify the capabilities of our method.

CCS Concepts: • **Human-centered computing → Visual analytics**;

Additional Key Words and Phrases: Social media, map, information diffusion

This work is funded by NSFC Key Project No. 61232012, the National Program on Key Basic Research Project (973 Program) No.2015CB352503 and the National Key Research and Development Program of China (2016QY02D0304). It is a continuous effort from previous funding of NSFC No. 61170204. This work is also supported by PKU-Qihoo Joint Data Visual Analytics Research Center. This work is partially supported by German Science Foundation (DFG) in priority research program SPP 1894 on Volunteered Geographic Information.
Authors' addresses: S. Chen, S. Chen, Z. Wang, and X. Yuan (corresponding author), School of EECS, Peking University, 5 Yiheyuan Road, Beijing, China; emails: simingchen3@gmail.com, {shuai.chen, zhenhuang.wang, xiaoru.yuan}@pku.edu.cn; J. Liang, Faculty of Engineer and Information Technology, The University of Technology, Sydney, Australia; email: christy.jie@gmail.com; Y. Wu, Southwest University of Science and Technology, China; email: wyd028@163.com.

2157-6904/2018/11-ART11 $15.00
https://doi.org/10.1145/3183347

ACM Reference format:
Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Yadong Wu, and Xiaoru Yuan. 2018. D-Map+: Interactive Visual Analysis and Exploration of Ego-centric and Event-centric Information Diffusion Patterns in Social Media. *ACM Trans. Intell. Syst. Technol.* 10, 1, Article 11 (November 2018), 26 pages.
https://doi.org/10.1145/3183347

1 INTRODUCTION

A large number of people use social media (e.g. Twitter, Flicker, Sina Weibo, etc.). Every day, millions and even billions of people from all over the world interact with each other, share opinions, follow and participate in significant events via posting, replying, or re-posting, producing a large number of messages that spread in social media platforms. The richness of social media data offers fruitful information and reflects social behaviors of people.

Analytical systems with data mining and artificial intelligence methods are effective for problem-solving in many areas. Visual analytics combining both machine and human intelligence makes such intelligent analytical process comprehensive [35, 49]. A visual analytics process can be modeled as a knowledge generation process involving data, human, and model [42]. In this manner, a visual analytics system is an intelligent system that supports user interactions and feedback. Social media data, which are dynamic, multidimensional and fuzzy, benefit from such human-involved visual analytics processes [16]. In this work, we propose a visual analytics method to analyze information diffusion processes in social media.

Existing visualization techniques mostly focus on illustrating how social objects (e.g., a message, a topic, or an opinion extracted from messages) spread over space and time [7, 57, 61]. Little research has focused on analyzing large amounts of related social object spread scenarios based on reposting behaviors. To clarify such information diffusion, we summarize two major scenarios: ego-centric diffusion and event-centric diffusion. Revealing how people get involved in the diffusion process and are influenced by a central user who initiates the process with multiple original microblogs is ego-centric diffusion analysis. Our previous work focuses on this [15]. Event-centric diffusion analysis seeks to identify how messages discussing social events are reposted to large groups of users and change over time. There are multiple weibo sources discussing a specific event. We analyze these sources as well as their reposting weibos to understand how the event information is diffused. In this work, we summarize the general information diffusion patterns in the social media and propose a consistent solution for analyzing them.

From one source microblog message, we can construct a multilevel reposting tree based on reposting behaviors. Both ego- and event-centric analysis are based on many reposting trees. How one original microblog gets reposted can be visualized with a node-link graph [41], but understanding many reposting trees and revealing the social interactions among influenced users requires mentally merging these trees, which can be challenging. Therefore, there is an urgent need for a clear and intuitive summarization of the diffusion process to illustrate the spread of messages across different groups of people.

There are challenges to designing a visualization fulfilling these requirements. First, social media data are usually very complex; they are heterogeneous, big, and dynamic, containing both structured and nonstructured data, thus making the summarization of information spreading structures among communities difficult. Second, capturing a user's influence requires an in-depth understanding of his or her social behaviors and a detailed analysis of the user's historical communication records. Such analysis is usually difficult as a user's behavior patterns are complicated in the real world and may change frequently, making capturing diffusion dynamics and revealing regular diffusion patterns challenging tasks. Third, the visualization of information diffusion processes and patterns among different groups of people within social events requires encoding multiple types of information, such as relationships between users, their roles, and the messages that they are involved with. It is challenging to reveal a clear skeleton of the event evolution within the above features. Meanwhile, it is important to avoid clutter, such as overlapping nodes and edge crossings in the visualization.

To address these challenges, we introduce D-Map, an interactive information diffusion map that can summarize the historical information of diffusion processes initiated by a central user in a social space context by considering the communities of influenced users [15]. In this article, we present an extension of the previous work and summarize the general diffusion patterns within a consistent visualization framework, namely, D-Map+. We produce the map based on hexagonal tessellation to reduce visual clutter by eliminating node overlaps. In our design, social media users are visualized as hex nodes with color and size encoding their behaviors and roles. These people are grouped into different regions as communities on the map based on their behaviors. D-Map aims to analyze both ego-centric user's information diffusion patterns and event-centric diffusion based on multiple ego-centric users. Our map can form the social profiling of central users and social events. In this way, both the central user's social influence and event stages are visually summarized.

In particular, this work makes the following contributions:

- **Visual Metaphor Design.** We introduce a novel dynamic information map design to reveal the dynamic patterns of how people are involved in diffusion processes. The techniques ensure a clear and intuitive visual representation of an aggregated information diffusion process. (**Section 4**)
- **General Diffusion Analysis.** We propose a general information diffusion analysis method in social media. Both ego-centric users' information diffusion and event-centric stage evolution can be interactively explored. (**Section 5**)
- **Visual Analytics System.** We develop a comprehensive visual analytics system (Figure 9) incorporating advanced community detection techniques and multiple coordinated visualization views. It provides a solution for understanding the influence of users and their social interactions in a diffusion process. We evaluate our system with data collected from Weibo, the biggest microblog platform in China, and reveal many interesting real-world patterns that have, to the best of our knowledge, never been visualized before (**Section 7**).

This article is structured as follows. Section 2 reviews related work. After introducing the data in Section 3, we present the D-Map+ metaphor in Section 4. We present the visual analytics procedure and technical details in Sections 5 and 6. We demonstrate the use of our tools in two case studies in Section 7. We evaluate the system in Section 8. Finally, we discuss the limitations and future work.

2 RELATED WORKS

2.1 Social Network Visualization

Social network analysis and visualization attract increasing attention due to the rapid development of online social media [16, 56]. The extensive studies on social network cover a broad range of topics, including community detection [21], role identification [33], and, most recently, information diffusion and influence anaysis [31, 44]. Visualization techniques play a major role in analyzing the social network [27–29]. Most of the existing techniques focus on capturing the structure of a social network, which is visualized using node-link diagrams [27], an adjacency matrix [28], or a combination of both methods [29]. However, none of the existing techniques is developed for producing a network map to illustrate the diffusion pathways among different people and across diverse communities. This is the focus of our article.

2.2 Information Diffusion Analysis and Visualization

Information diffusion has become an important research area in the domain of social media analysis in recent years [25]. Studies cover a wide range of topics including showing the evolution of topics [18], influence analysis [48], and visualizing and analyzing diffusion processes [7]. Many visual analytics techniques have been developed to help users better understand the diffusion process via interactive exploration and analysis. For example, Marcus et al. [36] introduced TweetInfo for a flexible aggregation of tweets from spatial, temporal, and event dimensions, thus supporting an accessible exploration of the event propagation process. Viegas et al. [53] introduced Google+ Ripples, which employed a hierarchical circular packing schema to illustrate re-sharing behaviors and the message-spreading process. Cao et al. [7] introduced Whisper, a flower-like visualization designed for monitoring the information diffusion of a given topic in real time. Ren et al. [41] proposed WeiboEvents which enabled flexible annotations of an information diffusion process based on crowdsourcing. All these techniques are successful designs illustrating the information diffusion process from different aspects, but none of them produces a static summarization of the diffusion process in the form of a map by which diffusion patterns can be revealed at a glance. Recently, more studies have focused on exploring collective topic or opinion diffusion dynamics [47, 57, 59]. Multiple visual analytics techniques have been developed to detect anomalous spreading of messages [61], opinions [10], and user accounts with suspicious behaviors [11]. These are visual analytics techniques that focus on problems in different application domains, instead of summarizing the diffusion process, so they differ from our work.

2.3 Ego-Centric and Event-Centric Dynamic Network Visualization

Researchers have proposed advanced visualization methods for the dynamic network [1]. Animation and small-multiples are two general approaches [3]. Recently, to reveal more insights on relationship evolution, researchers proposed timeline-based approaches for dynamic network visualization [2, 19, 52]. In the dynamic network, identifying the key players and their influence is another critical analysis task for understanding information propagation [48]. Classification and cluster analysis are widely used for role identification [40, 50]. These techniques group users into categories of different roles based on their behavior features. Cha et al. [14] measured a user's influence on Twitter based on the indegree, the number of retweets and mentions. These problems also attracted attention in the visualization field. In particular, an ego-centric view enables a closer look at individual behaviors, thus providing more detailed behavior patterns [8, 45, 58]. For example, Brandes et al. proposed a ripple metaphor to display the passing of time in the biography of a movie actor [5]. Shi et al. chose 1.5D form to embed the network along the time axis, revealing both the temporal and ego-network structures [45]. Cao et al. [8] developed Episogram that discussed the data model in ego-centric social interactions. Different from these techniques, D-Map+ introduces a novel diffusion map design that illustrates how people across various communities are influenced in a social network. The proposed method captures both topological and content information on an ego-centric social interaction network, producing a social portrait of the central user, which has not been done before. In addition, social event analysis has gained more focuses in recent years. Whisper [7] and WeiboEvents [41] analyzed one popular topic or key Weibo posting using visual analytics. However, a complex event includes multiple varying topics across multiple groups, which is the analysis target in this work. Other researchers used a river-based visual metaphor to reveal the dynamic evolution of an event [47, 57]. More recently, Wang et al. proposed river-based IdeaFlow to tackle the lead-lag patterns of multiple idea (i.e., generalized topics) diffusion [54].

These works showed great capability in analyzing the overall event stages. But the river-based metaphor restricts the representation of individual roles, relationships, and activities in one dimension. Our approach projects such information to a 2-dimensional space to reveal more details about event evolution. Moreover, by extending the ego-centric analysis of D-Map in our previous work [15], we integrate ego- and event-centric analysis in a unified model. Social media analysis urgently requires such general analysis of information diffusion. To our best knowledge, we are the first to address it.

2.4 Map-Oriented Graph Visualization

Maps were originally used in data with geographic coordinates [17]. Hexagon maps are frequently used for heat-map visualization for spatial temporal data [30, 55]. In this work, we mainly focus on using a map representation to visualize abstract network data. There are earlier works on representing network data with map-like visualizations. Guo et al. [26] used hexagonal maps to visualize the Self-Organizing Map (SOM). Gansner et al. [23] introduced GMap, an interactive visualization design that transformed a social network into a map view to highlight the boundaries between different communities. Furthermore, they proposed a stable layout of such map views for dynamic network data [31, 37]. They applied the dynamic maps generation techniques in Twitter data [24] and computer science literature [22]. Though these works have done an excellent job in preserving a human's mental map in analyzing the dynamic data, their focus is not the ego-centric users' social connections. Cao et al. [12] introduced FacetAtlas, based on a node-link graph visualization and bundling techniques, to represent a multifaceted atlas of a text corpus. Following a similar idea, Nachmanson et al. [38] introduced GraphMaps, which also applied edge bundling in a node-link diagram to help with the exploration of large graphs. Yang et al. [60] proposed a hexagon-tiling algorithm to visualize hierarchical data. Recently, Cao et al. [11] visualized a social interaction graph based on a triangle map for multidimensional data [9]. However, none of these techniques produces a compact visualization of the portrait of a centric social media user to illustrate his or her influence regarding spreading messages, which is the focus of this article.

3 DATA DESCRIPTION

In this study, the data are extracted from Sina Weibo, whose primary services are very similar to those of Twitter. Each weibo is a micro blog, as a tweet on Twitter. After one user posts a webio, another can add comments and repost the source weibo. The reposts can be reposted by other people again, which constructs a multilevel reposting tree. From a single weibo, we collect all the direct reposting weibos according to Sina Weibo API. Taking the source weibo as the first level, we can build the reposting trees level by level. The levels of a reposting tree vary due to many factors, such as the influence of participating users and the contents of the source weibo. To analyze the information diffusion pattern, we extract the weibo user, content, timestamp, id, and the pid, which is the id of the reposted weibo. A source weibo is the root of the reposting tree and its pid is null. We built a repost network of users by merging multiple source weibos with their reposting trees based on the same participating users. This network is modeled as a directed multi-edge graph $G = (V, E)$. The graph nodes V consist of all users in the resposting trees, and each edge $e = (u_a, u_b) \in E \subseteq VxV$ represents that user A reposts from user B one time.

How we collect these source weibos as target depends on the analysis scenarios. We summarize two important scenarios in social media: ego-centric analysis and event-centric analysis. When we collect all the source weibos from the same central user on his Weibo homepage, we can conduct ego-centric analysis of the diffusion patterns [15]. Moreover, by collecting multiple source weibos discussing posting and reposting about special events based on hashtags and keywords, we are able to perform event-centric analysis to identify information diffusion patterns in event evolution.

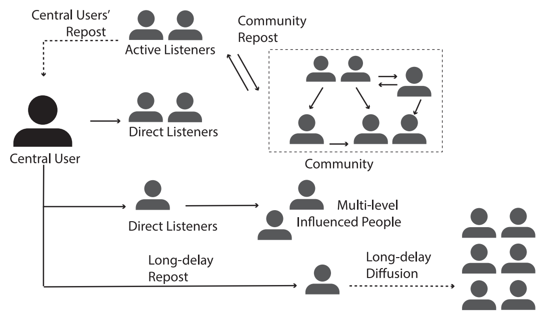


Fig. 1. An illustration of weibo data. People play different roles in one central user's reposting network with multiple behaviors.

These data construct the social network rooted in multiple users. First, we analyze the detailed features of information diffusion from ego-centric perspectives (Figure 1), which is the basis of event-centric analysis. Event-centric analysis is composed of multiple related ego-centric users discussing the same events. By checking these data, we summarized the characteristics of information diffusion in social media:

- (**D1:**) **Participating Features:** Central users attract different numbers of participants. Among them, active people repost weibos frequently, and inactive ones repost once.
- (**D2:**) **Influence of the Participants:** Participants' reposts lead to different times of multi-level reposts. The number of both direct reposts and total reposts that one user attracts indicates the impact.
- (**D3:**) **Key Player Distribution:** We can define people whose weibos are largely reposted as key players. Key players could have impacts on different groups or types of people.
- (**D4:**) **Dynamic Diffusion:** The life cycle of the diffusion of social media consists of multiple stages, including beginning, bursting, and dying. Reposting frequency, latency, impacts, and involved people are different in each stage.

In addition to the general features of information diffusion, there are special features we care about while analyzing ego- and event-centric information diffusion.

- (**D5:**) **Social Influence of Ego-centric Users:** How the weibos of central users diffuse across different groups of people represents their social influence. Different groups of people care about certain topics, which can be reflected in the diffusion process.
- (**D6:**) **Event Evolution of Social Events:** The number of important stages in a social event is a critical feature in event-centric analysis. Who plays important roles and how information diffusion affects event evolution are important features of social events.

Our design consideration is to explore the diffusion process and user relationships based on these features to produce a deeper understanding of users' social behaviors.

4 D-MAP+

In this section, we provide a conceptual model for designing D-Map+ and detail the visual design and construction process.

4.1 Conceptual Model

We aim to identify the general information patterns in the social network. Specifically, we are interested in how source weibos are diffused among multiple groups of people. In this process, key

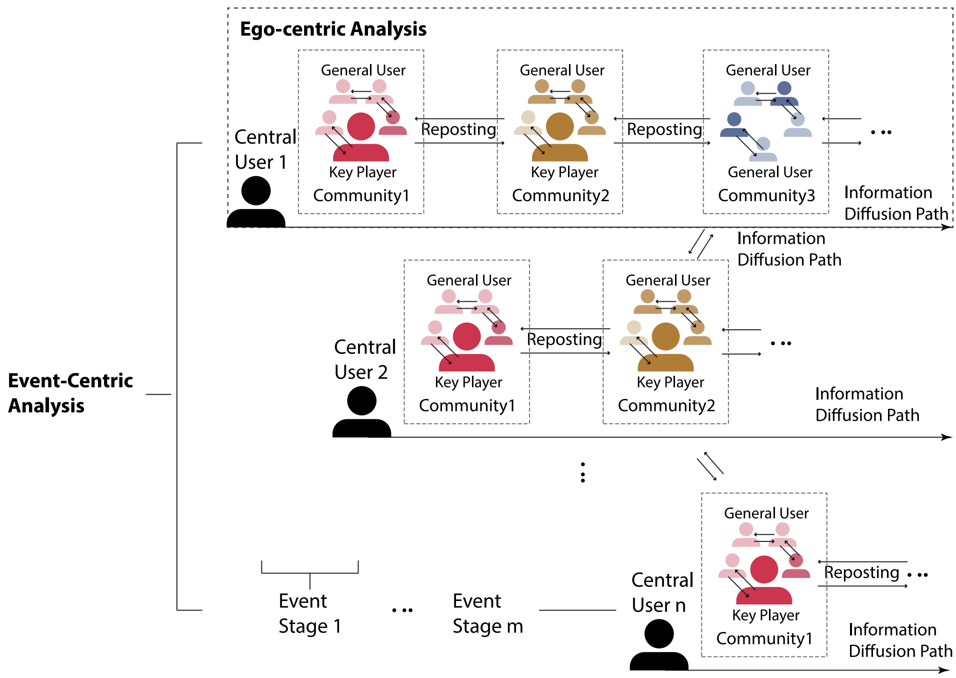


Fig. 2. Conceptual model illustrating the diffusion process. For ego-centric analysis, starting from a central user, information diffuses within and across multiple communities through a series of reposting behaviors. Collecting multiple source users with their repostings enables analysis of how events change in different stages, which produces event-centric diffusion analysis.

players and important diffusion paths should be pointed out. To further understand the semantics, summarized keywords should be provided. How people interact with each other (i.e., reposting weibos) could reflect the social structures of users as well as event evolution. Ego-centric information diffusion lays the foundation for the conceptual model [15]. By extending the ego-centric analysis with event-centric analysis, we summarize the general diffusion model in social media (Figure 2). The model emphasizes temporal information in event evolution. In different time periods, there could be key players initializing or reposting a message, which leads to large amounts of repostings. There could be multiple communities discussing in multiple event stages on the timeline.

To achieve these goals, we need to merge all the reposting weibos and conduct the analysis. A direct node-link graph visualization of communication records might be helpful, but it usually leads to a cluttered visualization that fails to reveal data insights effectively [46]. The hair-ball clutter prevents users from making sense of the group distribution and makes it hard to select individuals. The links add too much interference for analysis, and the visualization wastes space with a large amount of blank areas. Moreover, it lacks temporal information to investigate the diffusion process further. Thus, considering both the limitations of the force-directed graph and the characteristics of reposting behaviors (**D1-D6**), we summarize the design requirements.

- **G1: Show uncluttered participants' community distribution:** To investigate the features of participating people (**D1**), we need to categorize and group people with similar reposting behaviors. This forms the basis for investigating how ego-centric users influence multiple groups of users (**D5**) and users' reactions during events (**D6**).

- **G2: Understand social interactions among people:** Repostings lead to message diffusion, reflecting social interactions. Based on such interactions, we can infer users' social influence (**D5**). Thus, we need to compare users' reposting patterns (**D1**) and social influences (**D2**). Understanding how information diffuses in each event stage is essential for event evolution analysis (**D6**).
- **G3: Understand reposting features of key players:** Profilings of central users as well as social events are built with the reposting features (**D1**). As the representatives of all users, key players and their connections should be highlighted (**D3**). How key players act shapes the ego-centric users' network (**D5**) as well as event evolution (**D6**).
- **G4: Tell stories of dynamic diffusion:** Understanding the diffusion process requires understanding multiple diffusion stages (**D4**). We should allow users to select diffusion states and individual paths for details. Thus, users can explore the temporal aspects of ego-centric users' influences (**D5**) and event evolution (**D6**).

To fulfill the preceding requirements, we propose the D-Map+ design to project the information diffusion process onto a map with explorable features.

4.2 Community Detection

A community is a set of nodes which are densely and internally connected while having sparse connections with other sets. People who often repost the same individual's weibos and have similar behaviors can be considered as a community. As the basis of D-Map+ design, we need to detect communities of all participants based on their reposting behaviors. The input graph of the map is the multi-edge reposting network of social media users, merged from all the reposting trees of the source weibos (Figure 5(a)). The dashed link connecting nodes in different trees indicates that the connected nodes represent the same person. After the merging process, each node is one social media user, and each edge represents that user A reposts from user B one time. There could be multiple edges between two nodes. To find the community structure of the multi-edge graph, we use degree-corrected stochastic blockmodels [34]. The advantage of this method, which fits our design goal, is that it can not only identify the node's community assignment but also find the interactions between communities. As well, we do not rule out possibilities of using other algorithms. Let G be an undirected multi-edge graph on n nodes. It is assumed that there are K groups, and g_i is the group assignment of node i. Here, we give the unnormalized log-likelihood function:

$$L\left(G|g\right) = \sum_{rs} m_{rs} log \frac{m_{rs}}{k_r k_s}. \tag{1}$$

m_{rs} is the total number of edges between group r and group s. k_r, k_s are the sums of the degrees in group r and s, respectively. The goal is to maximize the probability on the group assignments of the nodes. The network is divided into an initial random set of k communities. By repeatedly moving a vertex from one group to another, the method will find a state with the highest score L.

K is determined when L is maximized. Following the works of Peixoto [39], we can set a minimal and maximal range for K calculation. For large groups of people (e.g., more than 10,000), we set the range of K from 5 to 30. In our test, most users' community results fell in this range. Users can also adjust the range in different scenarios.

4.3 Visual Encoding

To avoid clutter, we chose a compact layout, with the candidates of a mosaic map and voronoi-based tessellation. We chose the mosaic cartograms because they communicate data with countable integer units, which is easy for visual comparison [6]. We would like to choose a shape to minimize

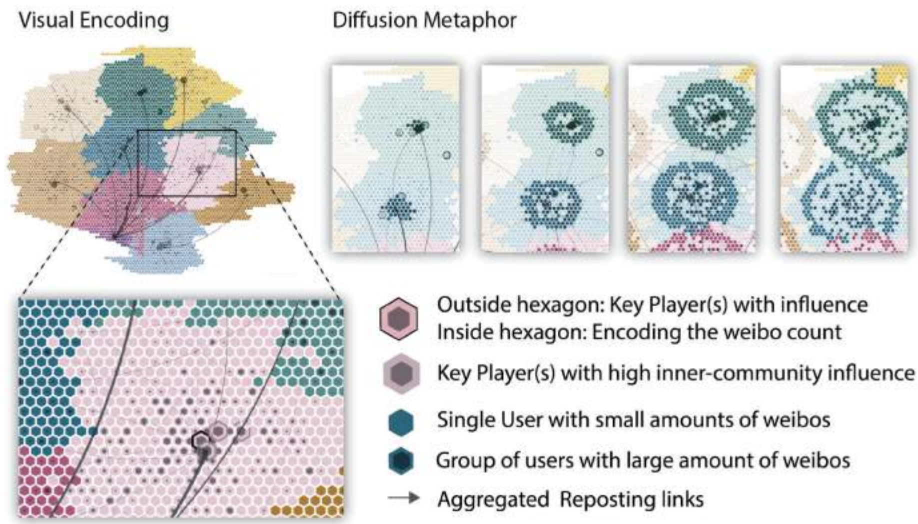


Fig. 3. Visual encodings of D-Map+. Each node represents a group of people participating in the diffusion process by reposting weibos. The color encodes the community, and the link summarizes the aggregated reposting directions.

wasted space among items and maximize the area inside them. The triangle grid introduces two types of triangles—regular and inverted—which may introduce ambiguities in visual representation. Square binning appears stretched out in the vertical and horizontal directions [13]. Other shapes with larger numbers of edges are too complex. The point and circle grid are not compact. Hexagons, which are widely used in the cartographic domain [30], are common in nature, further enhancing the aesthetic quality, familiarity, and acceptance from users [20]. Considering these factors, we finally chose to use a hexagon grid as the basis of D-Map+.

In the map design, each node represents one person or a group of people with similar behaviors. Each color region with multiple nodes indicates a community (Figure 3). The central user is drawn with a highlighting orange stroke. The key players are determined by a threshold value of the number of reposting people. In our experiment, we set the threshold as the square root of the total amount of people under consideration. The key players are highlighted with an enlarged hexagon with a black stroke to indicate they have a stronger influence on others. Within each hexagon, there is a smaller hexagon, the size of which shows how many weibos these people have reposted. To avoid clutter, we show the aggregated links among communities by default and show individual links of selected people on demand. The width of a link encodes the number of all repostings between two communities. The repostings include both direct and indirect repostings. Users can control a threshold to filter the numbers of repostings to reduce clutter. The nodes in each community are laid out from the inside out, based on their relative times, indicating the dynamic diffusion process for each community (Figure 3). There is a design tradeoff for such reordering. To gain awareness of critical temporal relationships, we may lose the topological relationships in the local cluster. To compensate for that, users can perceive the relationship through multiple interactions. Furthermore, users can still perceive the distances as relationships among different communities.

Region size represents the number of people in each community. We use color regions to encode different communities. For each community, we choose the corresponding ranked color in the color series of most representative feature (Figure 4). To compensate for lost details, we provide

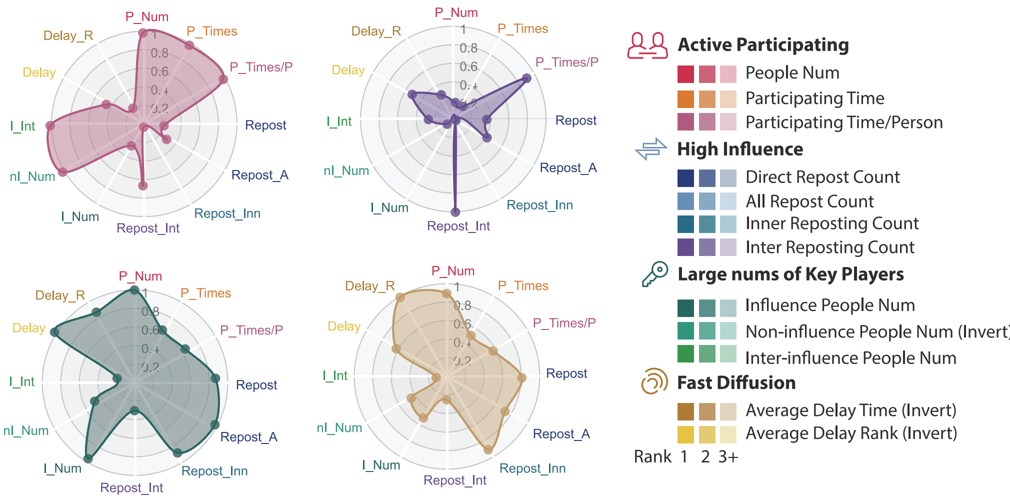


Fig. 4. Color decision for D-Map+. Four series of color encode the high-dimensional features of the community. The radar visualization shows the distribution of each dimension of selected communities.

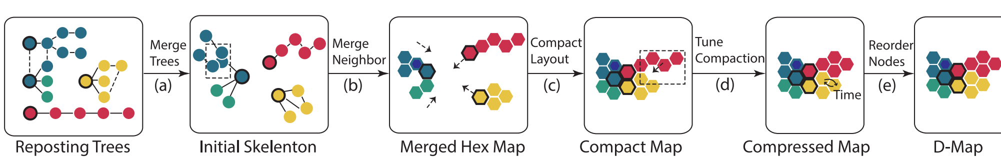


Fig. 5. D-Map+ construction process. With the input of a series of source weibos from one or multiple central users and the reposting trees of these source weibos, we can merge them into a graph, compact the layout, and reorder nodes in each community based on chronological order.

an interactive radar-style visualization to illustrate the normalized distribution of each dimension. Each axis is a subcategory of a feature, derived from data characteristics we discussed in Section 3. Each axis's name is the abbreviation of the corresponding subcategory, with the same color on the right (Figure 4). Thus, users can understand why each color is chosen. Other candidates of representative features are also perceived.

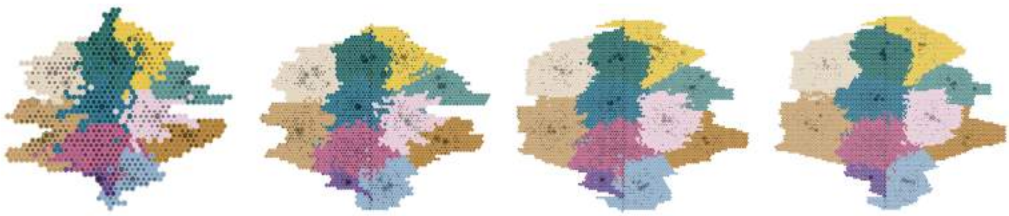
4.4 Map Construction

With the input of the multi-edge network with detected communities (Figure 5(a)), the map construction procedure includes customized force-directed layout, nodes merging, layout compacting, layout tuning, and reordering based on time (Algorithm 1).

To keep people in the same community positioned together, we chose a force-directed layout with customized link settings [32]. In addition to the original links between people, we add an artificial type of link in the graph. As illustrated earlier, we have a series of source weibos from a central user. We add edges among participants who repost the same source weibo. The edge-adding process makes people who repost the same weibo stay close, which may indicate that they share a similar interest. Moreover, it adds forces inside each community which contribute to better separating communities in the final D-Map+. In the next step, we merge the nodes within a distance threshold to reduce visual complexity (Figure 5(b)). These nodes usually have similar behaviors, so they are forced together. We apply a hierarchical merge operation in each community. We calculate the pair-wised distances of each node. After sorting the distance values, we start

ALGORITHM 1: D-Map+ Layout Algorithm

```
Require:
    A list of people nodes V_i with initialized force-directed layout position (step 1, omitted) V_i.pos, i = 1,
    2 ... n;
    Detected community number C;
    Expected output hexagon map size (hexagon number) S_h;
    Dividing parameter for compacting the layout N, making 360° into N pieces;
Ensure:
    A list of hexagon map points V_i, i = 1,
    2 ...m;
//Step 2: Merge the nodes that are close
Calculate the minX, maxX, minY and maxY of V
Collecting nodes from V into each community V_c, c = 1, 2,..C
Equally divide V_c into M blocks
for i = 0;i < M;i++ do
    Calculate the pair-wised Distance Matrix DisMatrix_c of block M_i
    MergeNum = (V_c.length * S_h)/(V.length * M)
    Sort nodes in M_i of V_c, and merge the nearest MergeNum nodes
end for
Merge all the nodes in M blocks of each community V_c and get S_h hexagons
//Step 3: Compacting hexagons into the center
Build N direction histograms histogramDir, each with direction range [h/360°, h + 1/360°], h = 0, 1, ...
N-1
Get the centerNode position
Push each hexagon V_i (i = 1, 2 ..m) into each histogramDir bin based on direction, sorting based on the
nearer distance to the centerNode
Initialize a candidate position queue; queue.enqueue(centerNode); count=0;
while queue.length > 0 and count < V.length do
    currentNode = queue.dequeue()
    for i = 0;i<currentNode.neighbors.length;i + + do
        if currentNode.neighbors[i].notOccupied then
            queue.enqueue(currentNode.neighbors[i])
            currentNode.neighbors[i].notOccupied = false
        end if
    end for
    dir = DIR(currentNode,centerNode)
    hexgon=histogramDir[dir].top();
    if hexgon! =NULL then
        histogramDir.pop(); Set position of hexgon as currentNode; count++
    end if
end while
//Step 4: Compact into rectangle
Move hexagons to the center horizontally and vertically (repeat step 3, N = 4).
//Step 5: Reorder based on time
for V_c in each V do
    Sort V_c based on the relative time of its source weibo
    Calculate the geometry center of nodes in V_c
    Map each hexagon V_ci from center out
end for
```



(a) (b) (c) (d)

Fig. 6. Different merging threshold settings for size control. (a) Set hex-bin number at 1,000. (b) Set hex-bin size equal to 3,000. (c) Set hex-bin size equal to 6,000. (d) Maximize the hex-bin size (9,000 in this case).

merging two nodes with least distance value. Through repeatedly merging, we obtain the merged nodes with the expected granularity of hexagons. Users can adjust the granularity by adapting to different scenarios (Figure 6). To reduce the computational complexity, we divide each community into multiple blocks equally and run the merging process for each block. Finally, we merge all the nodes in each block to get the final ones (Algorithm 1-Step 2).

After the merging process, we need to delete the blanks among nodes and make the layout compact (Figure 5(c)). Using force strength, we attract each node from different directions toward the geometrical center. Users can apply different dividing values of 360°. With an experienced value of 45°, we could attract nodes while keeping the relative positions of the neighborhood. To achieve the attraction process, we use eight direction histograms storing nodes in each 45° range and pop up the nearest nodes to the center one by one (Algorithm 1-Step 3). After the attraction, sometimes a large number of nodes are packed in a particular direction. To solve this problem, we apply a second-round compacting process to make the layout into a rectangle, which saves space and increases the data-ink ratio (Figure 5(d), Algorithm 1-Step 4). In each community, we calculate the relative time of each weibo compared with its source weibo. We set the minimal time for the node if it contains multiple weibos. We calculate the center of each community and map the nodes from the inside out according to their relative times (Figure 5(e), Algorithm 1-Step 5).

4.5 Semantic Understanding

The D-Map+ provides keyword layers over the map. It calculates the specific users' topics of interest and visualizes them with a word projection on the map (Figure 9). With the projection of these keywords, users can easily understand and follow the main discussion themes and investigate the information diffusion process. The size of the keywords encodes the frequency of mention. Users can turn the keywords layer on the map, depending on different tasks.

4.6 Interactive Exploration

The D-Map+ is a dynamic visualization, with multiple user-friendly interactions. The interactions are designed so that analysts can explore participants' distributions and relationships, as well as the diffusion process in social media. We divide the supported interactions into the following categories.

- **Navigation and Semantic Zooming:** The basic functions of spatial temporal exploration for a map are panning and zooming, which are supported in D-Map+. Besides geometric zooming (Figure 8(b)), we also provide semantic zooming functions for users to expand the detailed weibos. Users can expand the community of interest for detailed exploration (Figure 8(c)). In the zoomed map, each hexagon node corresponds to one user, which shows more details without aggregation.

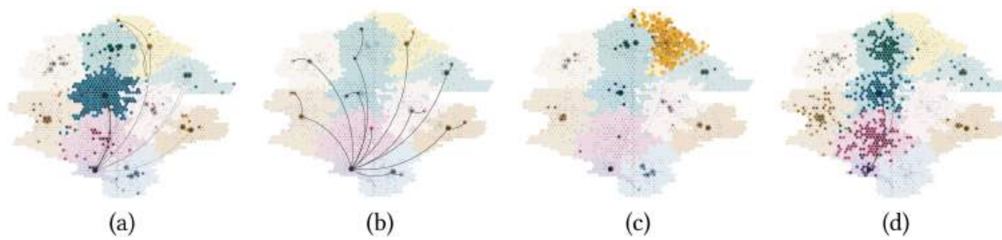


Fig. 7. Interactions on the Diffuse Map. (a) Selecting nodes' direct reposting. (b) Select multiple people from different communities to investigate the diffusion paths. (c) Select key players and see how they influence the exact community. (d) Following specified events and check the overall diffusion.

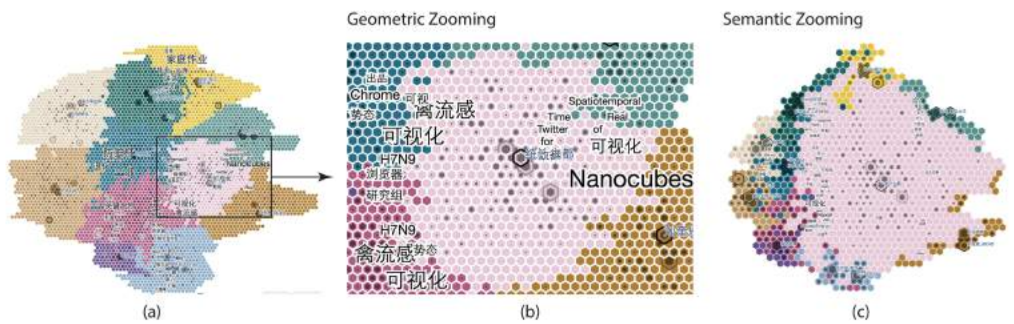


Fig. 8. Two zooming mechanisms. In the original map (a), users can use the mouse wheel to conduct geometric zooming (b) and click the targeted community for semantic zooming (c).

- **Node Highlighting and Selection:** We support single and multiple node selections. Single selection is provided for fast-forwarding different people's behavior (Figure 7(a),(c)). Multiple nodes selection enables users to select multiple nodes as a group for further information extraction (Figure 7(b)). We also enable users to select both nodes and communities.
- **Information Extraction:** After selecting, details of the statistics of selected nodes can be derived. The information is also divided into four categories of communities and shows the number and ratio of selected results (Figure 4). Keywords of selected weibos will be laid on the map for semantic exploration. Hovering on hexagons will pop up the detailed raw weibo contents.
- **Diffusion Exploration:** As the hierarchical relationship is flattened in the map, we provide interactions to identify the reposting relationship. Users can explore the direct post and repost relationship from multiple events (Figure 7(d)). If one user leads a large reposting, how information reaches him and diffusese from him are critical and can be explored interactively.
- **Linked Highlighting:** We support multiple linking functions by highlighting the nodes in other linked visualizations.

In short, D-Map+ is a customized visualization that represents participation in social communities and describes the diffusion process. To enhance the analytical capability of D-Map+ from multiple aspects, we propose an interactive visual analytics system.

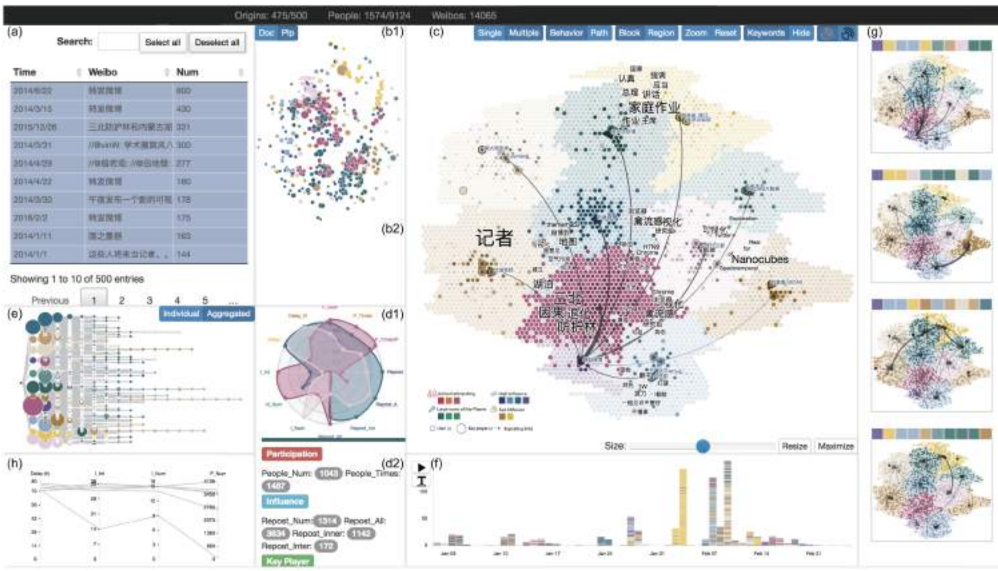


Fig. 9. System Interface. Source Weibo Table View (a), for selecting different groups of source weibos; Source Weibo Distribution View (b), including Documents View (b1) and Keywords View (b2); D-Map+ View (c), summarizing the social interaction among participating people of a central user; Community Radar View (d), showing the high dimensional features of communities with a Radar View (d1) and a Statistics Information Window (d2); Hierarchical View (e), illustrating the reposting structures; Timeline View (f), highlighting the temporal trends of the diffusion; Small Multiple View (g), identifying key time frames of D-Map+'s snapshots; Parallel Coordinates View (h), showing features of diffusion paths.

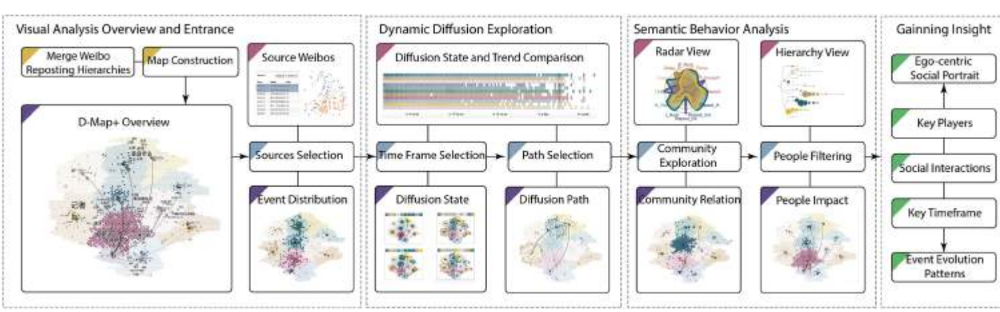


Fig. 10. Visual analysis pipeline illustrating how to explore users' D-Map+ with the augmented trend, hierarchy, and high-dimensional analysis. Diffusion path, social interactions, and people impact can be found. Besides the general diffusion patterns, people profiling for ego-centric analysis and event evolution for event-centric analysis can be achieved.

5 VISUAL ANALYTICS PROCEDURE

The visual analytics system combines D-Map+, Souce Weibo Table View, Community Radar View, Hierarchical View, Timeline View, and Small Multiple View (Figure 9). By analyzing multiple aspects of weibo data, users can explore the diffusion process among communities systematically (Figure 10). The color is coherent in the system and mapped to the detected community (Figure 4).

Our analytical process follows our previous taxonomies for social media visual analytics [16]. We have six goals in four stages in the pipeline, including visual monitoring, feature

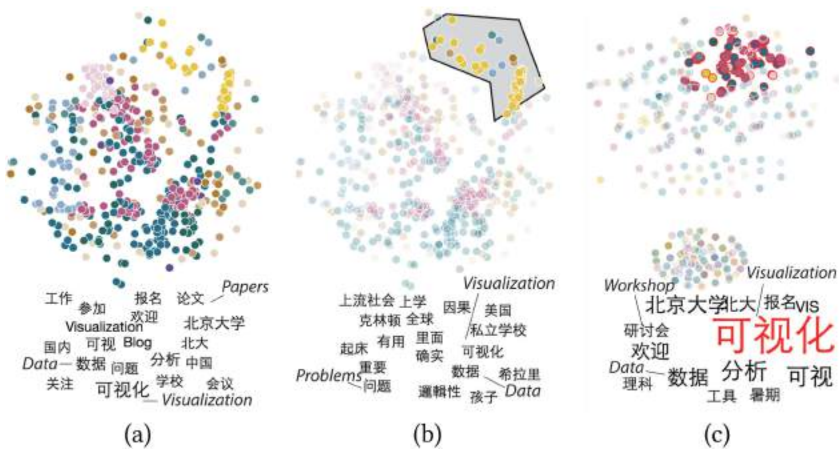


Fig. 11. Weibo sources analysis with (a) reposting people distribution distance and (c) document distance. Interactions, including (b) brushing and (c) keyword filtering, are supported.

extraction/event detection, outlier detection, and predictive analysis/situation awareness. Our proposed analytical process (Figure 10) follows these stages. First, our map design provides dynamic overview and animation to support visual monitoring (**Overview and Entrance)**. Based on this observation, we extract diffusion features and detect diffusion stages interactively. Later, users can detect special behaviors, such as bursting reposting behaviors, complex reposting structures, and the like (**Dynamic Diffusion Exploration**). Finally, we allow users to gain semantic information for generating insight, which supports the situation awareness process (**Semantic Behavior Analysis**).

5.1 Visual Analysis Overview and Entrance

By projecting the source weibos from a central user into a 2D space, we provide a starting point for analysis. One of the research goals is to understand the characteristics of communities and which types of information some communities prefer. There are two key points: participating people and weibo content. On the one hand, users can analyze people relationships, which are reflected by the distributions of participants in reposting each source weibo. On the other hand, users can explore the participants' preference for different keywords and content to further understand the characteristics of communities. Therefore, we enable users to analyze weibo sources from these two perspectives. By default, we construct a high-dimension vector for each source weibo. Each dimension is the people count of each community. Consistent with Section 4.3, we chose the color of the community with the largest participating number to encode the source weibo. With the calculated high-dimensional distance, we project the documents into a 2D space with t-SNE [51] (Figure 11(a)). From the content perspective, we first process the text of the source weibo by word segmentation and remove the stop words. Stop words include standard terms without specific meaning. To get the distance matrix, we adopt Term Frequency-Inverse Document Frequency (TF-IDF) [43] to create a weighted vector and measure the similarity of each source weibo based on the cosine distance between the vectors. Finally, we project the source weibos to the 2D space based on their content similarity with t-SNE (Figure 11(c)).

Interactions such as clicking and brushing selection (Figure 11(b)) are supported, and users can also click on keywords to select related weibos and reposting people (Figure 11(c)). Moreover, we provide a table view of source weibos with sorting, keyword searching, and filtering functions

(Figure 9(a)). The selected source weibos will pop up for highlighting. Participants of the selected source weibos are highlighted on D-Map+ for further exploration (Figure 7(d)).

5.2 Dynamic Diffusion Exploration

We apply a Timeline View (Figure 9(f)) with a Small Multiple View (Figure 9(g)) to support the exploration of the dynamic diffusion process with D-Map+. There are two types of temporal representation in the Timeline View, including the aligned timeline (Figure 12) and the absolute timeline (Figure 9(f)). The absolute timeline shows the normal case of original time distribution about the axis. Considering the ego-centric scenario, sometimes we care about the general diffusion patterns, so alignment of the reposting weibos based on each source weibo provides the baseline for comparison and aggregation. In such setting, the y-axis is the detected community and the x-axis is the aligned relative timeline. Considering the short life span of each weibo, we show the first 24 hours of the reposting weibos with 80% of the time line width. We provide an animation function to fast-forward the diffusion process. We propose two methods to split the timeline and show the critical period range in small multiples, based on percentile division and entropy-based division. We can use the percentile group to produce an overview of information diffusion among communities (Figure 9(g)). We split the data with 25%, 50%, and 75% amount thresholds. For entropy-based division, we use Shannon entropy, which measures the distribution's degree of dispersal or concentration of communities. For a given histogram $X = \{n_i, i = 1, \ldots, N\}$, community i occurs n_i times in the sample. $S = \sum_{i=1}^{n}(n_i)$ is the total number of the community observations. H(X) is defined as the following:

$$H(X) = -\sum_{i=1}^{n}(n_i/S)log_2(n_i/S). \quad (2)$$

We aim to find time periods with low entropy values and large entropy changes. These are likely to have a concentrated community distribution of repostings inside the community, with low entropy values. The change of entropy indicates that the source weibos are diversely reposted to others from few communities or vice versa. In the Small Multiple View, key players are shown as rectangles in the order of influenced people number (Figure 9(g)). This provides good suggestions for identifying important event stages. When we click the thumbnail, the corresponding D-Map+ is shown in the main window. On the selected D-Map+, users can explore the particular diffusion paths of different people (Figure 7(b)). People in the highlighted key players' diffusion path can be identified as important participants. With these interactions, users can identify the key time frames and key players.

To provide additional statistics about the diffusion network's properties, we propose the parallel Coordinates View (Figure 9(h)). It shows the Delay Time (Delay), Inter-influence People Number (I_Int), Influence People Number (I_Num), and People Num (P_Num) of different diffusion paths on the map. These features are consistent with those of the community. Clicking a hexagon on the map to investigate how the information propagates to the corresponding users, a polyline will be drawn in the parallel coordinates to help users understand the diffusion process.

5.3 Social Community Analysis

A series of reposting behaviors lead to information diffusion, which reflects social interactions. Specifically, our system supports the investigation of the characteristics of each community, inter-community diffusion, and people impact.

First, in Community Radar View (Figure 9(d1)), the high-dimensional features reflect the community characteristics. When users select nodes inside a community, the selected people number will be shown (Figure 9(d2)). In addition to the statistics, an overview of inner community behaviors can be perceived through the arrow glyph design. These behaviors usually include single center diffusion and strong connections among community members. In addition, we can investigate the keywords distribution of weibos from different communities (Figure 9(c)). When users identify a specific groups of users, their aggregated keywords distribution will be distributed around them on the map. This helps users see the semantic information distribution of the weibos. Second, by selecting a community on the map, the correlated communities are highlighted (Figure 7(a),(c)). Thus, we can infer how much influence the community has and how diverse its users' influence is. Also, the Hierarchical View aggregates nodes of the same communities in the diffusion process, which helps users understand the position of a selected community in the hierarchical reposting tree. Moreover, when users select multiple communities, community features can be compared interactively in the Community Radar View. Third, by selecting the nodes on the map, we can investigate people's direct reposting and reposted nodes (Figure 7(a),(b)). Diffusion path and key players can reflect the central users' impact (Figure 9(c)).

6 SYSTEM IMPLEMENTATION

The visual analytics system is built with HTML5/Javascript, and the server-side processing is with Python and MongoDB. The client uses SVG with D3.js [4]. By implementing a customized message protocol, multiple views are linked in the client visualization. We crawled weibo data through the open APIs by Sina Weibo and constructed the reposting tree for each source data with Weibo Events Crawler [41]. In our system, we support users to submit querying keywords and central users' id offline. The data are stored in MongoDB and provided to the client using a customized API for fetching data.

7 CASE STUDY

We present two cases showing the ego-centric and event-centric analysis, respectively. There are two more cases discussing about the ego-centric profiling and community analysis with D-Map [15].

7.1 Case 1: Dynamic Diffusion Pattern Analysis

In this case, we explored the diffusion patterns among communities. We selected 300 weibos of one influential person and constructed a D-Map+ with 7,694 reposting weibos from 5,917 unique users (Figure 12). There were two large groups, with 2,986 (C1) and 1,811 people (C4), shown in red. By exploring the diffusion process, we can better understand how these communities formed and what their behavior patterns were.

There were two main diffusion states (Figure 12). The first state included three stages (T1 - T3). In the first 15 minutes (T1), the central user posted weibos and mainly affected the direct listeners group C1. Later, in one hour (T2), people in the surrounding communities reposted more weibos, while the weibos kept spreading inside C1. By selecting C2 (Figure 13-2) in T3, we found that it had the most inner spreading counts at 298. This indicated that people in C2 were active. In the later stage (in 10 hours), the reposts lasted and spread mainly within each community. Afterward, this transited to the second main state, which was also segmented into three important stages (T4 - T6). The purple community, C3, with high influence reposted weibos from C1 and shortly had burst diffusions in C3 and C4 (T5). Then, the information spread in all the communities (T6).

By further investigating the community property, we infer reasons for the communities' separation and behaviors. In addition to the large common first-listeners (C1, Figure 13-1), the central user had another long-delayed reposting community (C3). By clicking the key player in C3 (Figure 13-3), we found he was one of the most influential persons on Sina Weibo, one who had much

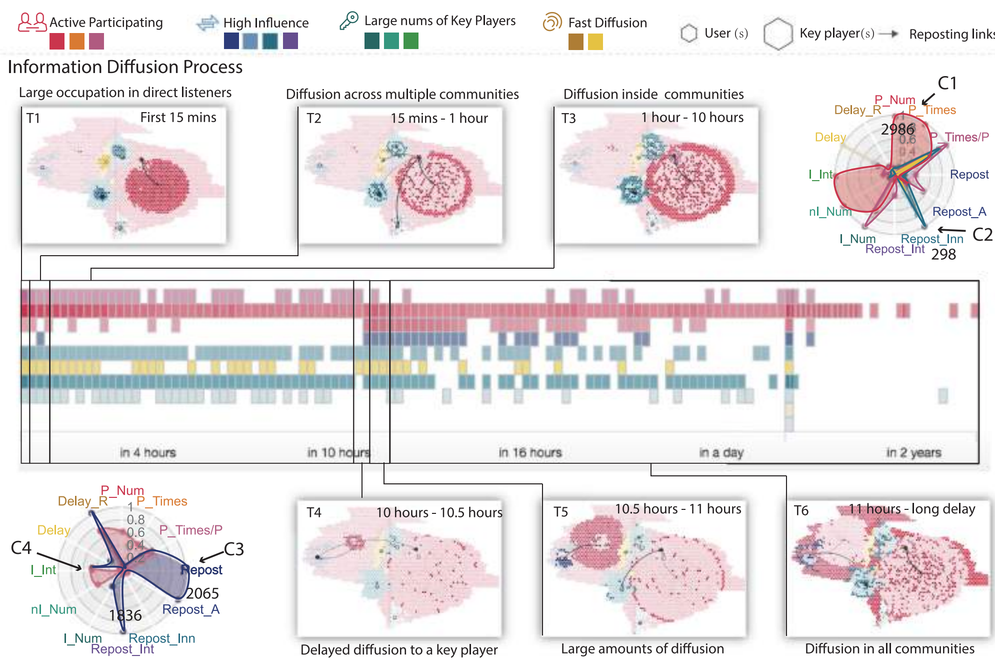


Fig. 12. Dynamic diffusion pattern analysis. Multiple diffusion stages are shown in the small multiples and the timeline. The first main status (T1 - T3) covers the right regions on the map. The second main status covers another reposting burstings in 9 hours (T4 - T6).

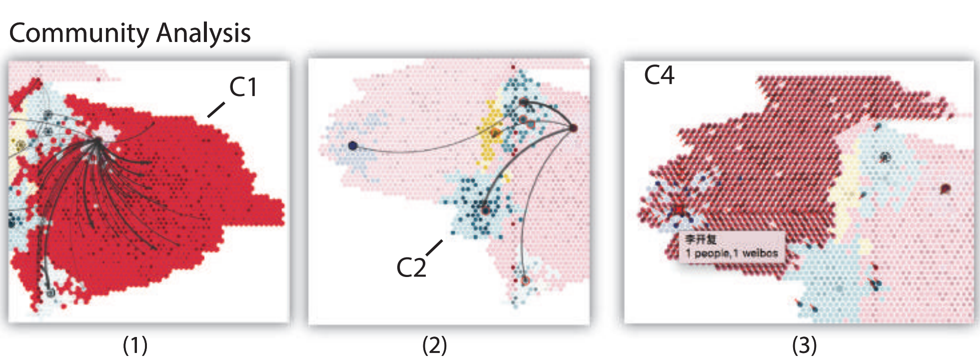


Fig. 13. Community analysis. By interactively exploring communities, we find key players and summarize diffusion patterns. One of the most influential people in Sina Weibo forwarding the root users' weibo only once could build up half of the social map.

more followers than the central user. Thus, we can be aware of the prominent levels of different people, as well as the diffusion state changes over time.

7.2 Case 2: Social Event Visual Analytics

In this case, we evaluate the capability of event analysis using D-Map+. We gathered data about a real-world social event that lasts for 1 month. There are 6,963 people with 9,626 weibos talking about this event. The start of the event occurred when a woman was left with a piece of gauze

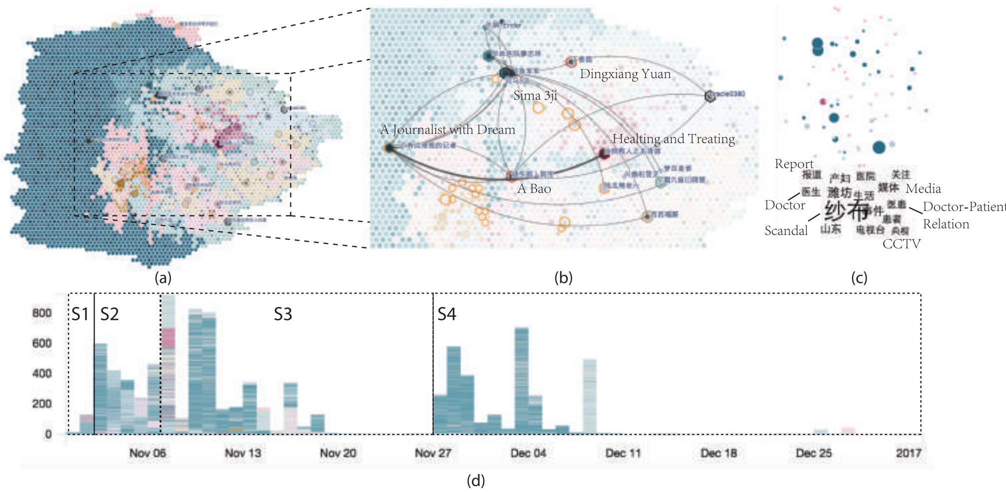


Fig. 14. "Gauze Scandal" social event overview. (a) Event map view of all the people we collected participating in posting and reposting the messages about the event. (b) Key players in the event. (c) Critical source events and keywords distribution.

insider her after surgery in a hospital in August 2016. A local TV program denounced the hospital and doctors for this on October 30. Initially, many people who watched the program were angry about the hospital and the doctor, as well as sympathetic to the victim. However, later, the program and local TV journalists were found to be unfair and didn't reflect the truth. The whole event experienced several stages with multiple perspectives, which were complex for the public to understand. We used D-Map+ to analyze the event and determined the key stages, key players, and what was going on in each stage.

We projected the 6,963 people onto the map (Figure 14). There are several major communities, one of which covered a half region of the map. From the summarized keywords overview (Figure 14(c)), we can see keywords including *gauze, media, hospitals, doctor, doctor-patient relationship, CCTV, TV*. With these keywords, we can have an overall filling for the events as starting points. The time distribution is shown in Figure 14(d), which covers the range of October 30, 2016 to December 31, 2016. By interactively exploring the key players with top reposting weibos, we found the key player distribution as well as their relationships (Figure 14(b)).

By identifying the key players and their active time ranges, we can further drill down to see their roles and the event stages in details (Figure 15). By identifying the weibos at the beginning of the event (Figure 14(d)), we noticed that the user "Junjun" was one of the key players. He was one of the journalists at the local TV station. All his weibos claimed just treatment by his local TV station and argued that the patient had experienced unfair treatment in the hospital. Some key players like "Sima" reposted weibos to support his claims. However, most of the people who reposted his weibos represented an unsupported attitude and didn't believe what he said (Figure 15(a1)). We can infer these attitudes from the summarized reposting keywords distribution: *chicanery, shameless, poor*, etc. Even the "Sima" who actively posted and supported "Junjun" had reposting using the same types of words (Figure 15(a2)). However, "Sima" also fought back by reposting them again, thus building up long reposting chains (Figure 15(a3)). This was the first stage of the event.

By extending the time forward, we could identify an important key player called "A journalist with a dream" (We called him "A journalist" later). He was a "We-Media" journalist with special

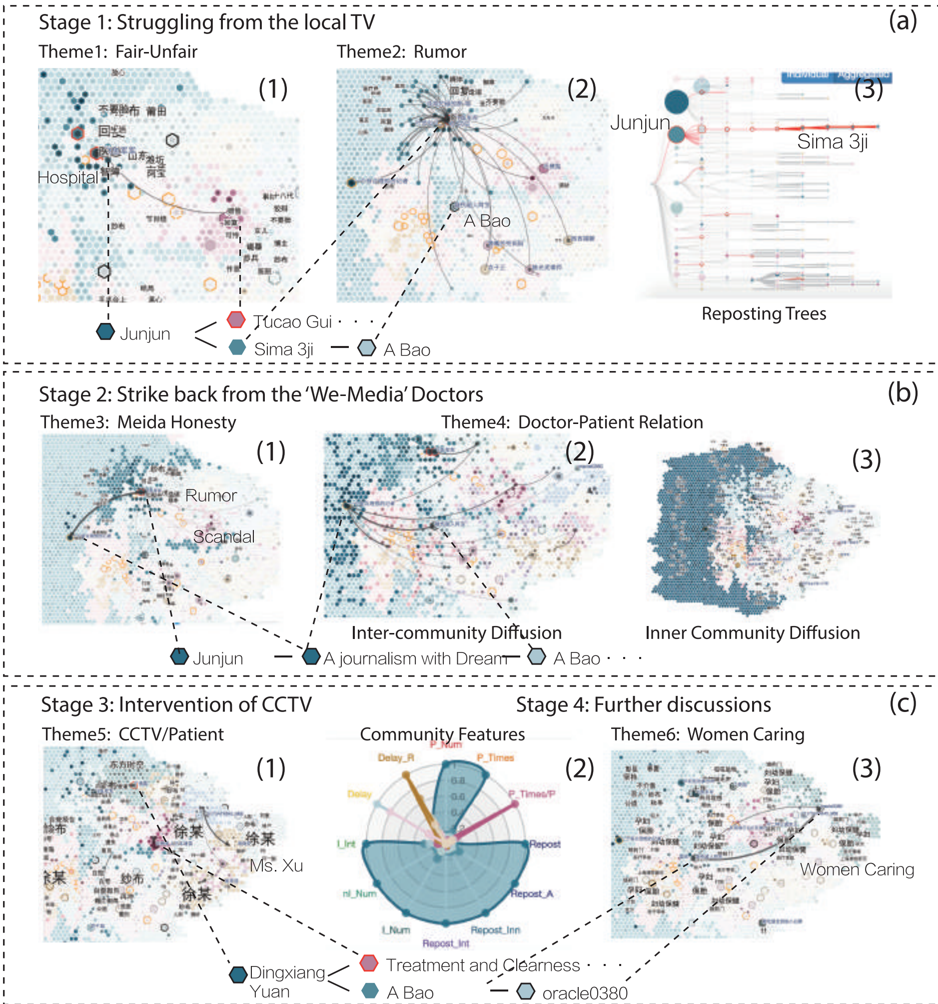


Fig. 15. "Gauze Scandal" social event stage analysis. There are four main stages that we investigated. (a) Stage 1 - Struggling from local TV. The key players includes "Junjun," a journalist from a local TV stations; "Sima," a supporter of "Junjun';' and other people who are against "Junjun" (e.g., "Doctor A-bao"). Stage 2 - Fight back from the Doctor "We-Media". The key player includes "Junjun," "A journalist with a dream," and "Doctor A-bao." Stage 3 - Intervention of CCTV. The key player includes "Dingxiang Yuan." Stage 4 - Further summary of the events.

interest in doctor-patient relationships. He reposted several weibos of "Junjun," and this led to large, multilevel repostings (Figure 15(b1)). He actively reposted materials, including interviews with the police, a summary of the current event stages, and more. We could see his weibos affecting multiple intercommunity repostings (Figure 15(b2)) and inner-community repostings (Figure 15(b3)). Among them, many "We Media" doctors such as "A Bao", built up a "We Media" union to argue for the fairness of the local hospital and the surgeon. They continuously searched for evidence and convinced the public. Since the participation of "A journalist" on November 1, 2016, the event was continuously exposed to the public over the following days, which we concluded as the second stage. On November 7, we found a peak in the time line (Figure 14(d)), and we checked the new discussion keywords and participants. In addition to the general existing keywords, "Miss Xu" and "Oriental Horizon" were two newly emerging ones. Many accounts mentioned that *"The Oriental Horizon of CCTV investigated the events and the so-called victim Miss Xu lied to the public."* Among them, posts from an online doctor service account "Dingxiang Yuan" received great attention and were reposted several hundred times (Figure 15(c1)). We could also observe an interesting feature in that several fast reacting accounts were active in these periods, indicated by their yellow color (Figure 15(c2)). Also, we found a detected community with a key player called "Treatment and Clearness" reposted the most (171 times) of all participants. Though he didn't lead to many reposts, this revealed his activeness in this event. He mainly discussed the details of the surgery and argued with others. Later, the so-called victim left the hospital and the discussion faded out gradually. In the last stage, only a few people summarized and discussed other points, such as women's healthcare issues in China.

With our methods, we could conveniently identify the event evolution and information diffusion from this real-world event. Key players, event stages, and main themes were visually captured. We also supported detailed explorations with interactions. By detecting the main diffusion paths and behaviors, we could summarize and further understand the key players' relationships (Figure 14(b)).

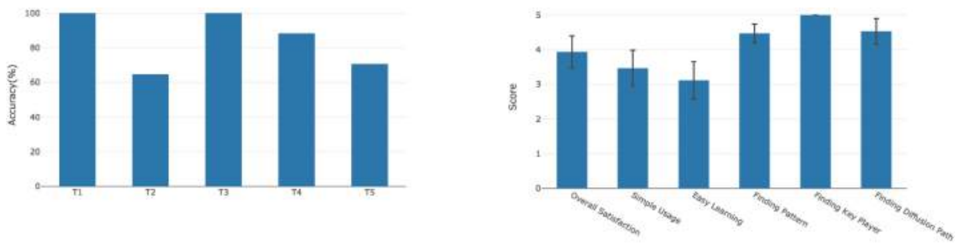
8 EVALUATION

In this section, we provide two aspects of our evaluations, including a user study with 17 participants and a deep interview with a sociologist.

8.1 User Study

To evaluate D-Map+, we performed a user study with 17 participants (11 male and 6 female, aged 18–30) with basic visualization knowledge. The goal of our user study was to evaluate whether our system fits the design goals (G1-G4). After providing a 10-minute tutorial, our participants were required to analyze real social media data from the case in Section 7.1. We recorded their answers and completion time for each task, and we asked them to complete a questionnaire. The user study tasks were:

- **T1: How many communities are there in this social network?** In the map design, each color region with multiple nodes indicates a community. Participants can get the number of communities from the color regions. **(G1)**
- **T2: Can you find which community has the most participants, give the representative color, and indicate how many people are there?** Participants should evaluate the participating features of different groups from the statistics view. **(G1, G2)**
- **T3: Can you find a community, name the key player(s) in the community, and describe the representative features of this community?** Key player(s) are encoded by larger size hexagons. Participants can find the key players on the map and analyze the community using the radar view. **(G2, G3)**
- **T4: Do you observe any temporal patterns in the reposting behavior of the social network?** The timeline view highlights the temporal trends of the diffusion. Participants should analyze temporal patterns by playing the animation or brushing the timeline to identify events. **(G4)**
- **T5: Who is the person leading to large amounts of repostings, what's his name, and what's the discussion about?** Following T4, this task requires participants to identify the key players in the reposting social network. **(G3, G4)**



(a) (b)

Fig. 16. User study results analysis. (a) Task accuracy summary. (b) Feedback of the questionnaire.

In summary, these tasks were designed to analyze if participants can understand the construction of the diffuse map and get information diffusion insight through the supporting views. In our observation, participants took 19 minutes on average to finish these tasks. After finishing the tasks, participants were required to rate the usefulness of the system.

Figure 16(a) shows the accuracy of the five tasks. All participants provided correct answers to T1 and T3, while the accuracy of T2, T4, and T5 were 64%, 88%, and 71%, respectively. After analyzing their interaction histories, we identified that some participants didn't use the radar view to analyze features of communities, which led to a wrong answer for T2. For T4, some participants didn't identify the two main diffusion states because they just analyzed some source weibos instead of selecting all of them. For T5, some participants found the key player in C3 but they didn't think he led a large amount of repostings.

Figure 16(b) shows the grades of six aspects evaluated by participants. In general, participants were satisfied with the system and thought it was useful to analyze diffuse patterns on social media. However, the rating for "Easy Learning" was lower than for others. This may be because the system interactions were complicated for these users.

Along with the user study, users were interviewed about the pros and cons of the system. Most participants felt that the map design was creative and the animation was helpful for them to have an overview of the information diffusion process. One participant stated, *"The map was elegant and saved much space in showing the information diffusion."* In addition, the participants also provided many reasonable suggestions from improvement. *"I felt the interactions were complicated and hard to start a investigation." "I couldn't get detailed info about one user in the reposting network."* These can be good resources for proposed improvements in the future.

8.2 Expert Feedback

We presented the system interface and results of the case study to a sociologist. In our discussion, she expressed great interest in what our system offered. In particular, she was impressed by the different impacts of the social roles of these central users on information diffusion patterns. She indicated that the results of these case studies inspired her to consider a research project on the relationship between the identity of a micro blog user and the consumption of the micro blogs of that user by the public. She believed that using our system can help her find the patterns of people's behaviors and that such information "*would provide more hints to generate some research hypotheses, such as whether information diffusion patterns differ between micro blogs from business sales persons and political advocates.*"

In short, we generally got positive results from the evaluations. We will improve the system to provide a neater and simpler visual analytics interface, as well as provide detailed exploration features.

9 DISCUSSION

The pros and cons of the visual design of D-Map+ has been discussed in a previous work [15]. From the perspective of ego-centric analysis, the visualization forms a diffusion map that portrays a user's social behaviors and reveals his or her influence regarding spreading information in the social space. For the event-centric analysis, the visualization allows users to explore dynamic event evolution with sufficient details and clear representation. This visualization enables a dynamic exploration of the historical diffusion processes and facilitates a fast comparison of diffusion patterns. In this discussion, we mainly consider the event-centered analysis and the coherent model integrating both ego- and event-centric analyses.

Event analysis requires the understanding of multiple event stages. One important goal is identify the key players and their messages, which pushes forward the event evolution. Our map design with dynamic visualization is suitable for such cases. In one aspect, we can make use of 2-dimensional space to illustrate peoples' relationships while using dynamic updating and interactive selection to support temporal analysis.

Compared with the existing river-based visual metaphor, our map generation can provide sufficient details of key players, diffusion processes, and semantic information for diffusion analysis. As illustrated in the last case, we are able to identify multiple event stages interactively. However, due to the aggregation of people, the temporal reposting patterns are somewhat hidden. To better identify the different roles people play in different time stages, we can extract individual weibos from the same persons and construct a more detailed map in the future.

Ego- and event-centric approaches both target information diffusion analysis. The difference is whether it is single central user analysis or multiple central users analysis. Our design integrates the two tasks into a unified model for exploration. Though successfully integrating the models, we also need to pay attention to the differences in these two tasks in our future design. Ego-centered analysis addresses relative time analysis, summarizing habits or regular patterns. Event analysis focuses on absolute time evolution and changes. Thus, specific temporal pattern mining features can be added into models to support more complex tasks.

10 CONCLUSION

We proposed a novel visualization method, D-Map+, to visually summarize and explore general information diffusion patterns in social networks. We map all the people reposting one or multiple central users' weibos to a hexagon map. Diffusion patterns and community interactions can be detected with a focus on key players and important diffusion paths. Semantic information could be derived with interactive exploration as well as the hints from mining results. With a comprehensive visual analytics system, we evaluated our work with real-world social media data and found interesting patterns to help us understand the unique features of both ego-centric and event-centric information diffusion.

ACKNOWLEDGMENTS

The authors wish to thank Xiaolong Zhang and the anonymous reviewers for their valuable suggestions and comments.

REFERENCES

[1] Daniel Archambault, James Abello, Jessie Kennedy, Stephen Kobourov, Kwan-Liu Ma, Silvia Miksch, Chris Muelder, and Alexandru C. Telea. 2014. *Temporal Multivariate Networks*. Springer International Publishing, 151–174. DOI: https://doi.org/10.1007/978-3-319-06793-3_8

[2] Dustin Lockhart Arendt and Leslie M. Blaha. 2014. SVEN: Informative visual representation of complex dynamic structure. *CoRR* abs/1412.6706 (2014). http://arxiv.org/abs/1412.6706

[3] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. 2016. A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum* (2016).

[4] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[5] Ulrik Brandes, Martin Hoefer, and Christian Pich. 2006. Affiliation dynamics with an application to movie-actor biographies. In *EuroVis06: Joint Eurographics - IEEE VGTC Symposium on Visualization*, Lisbon. 179–186. DOI: https://doi.org/10.2312/VisSym/EuroVis06/179-186

[6] Rafael G. Cano, Kevin Buchin, Thom Castermans, Astrid Pieterse, Willem Sonke, and Bettina Speckmann. 2015. Mosaic drawings and cartograms. *Computer Graphics Forum* 34, 3 (2015), 361–370.

[7] Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. 2012. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2649–2658. DOI: https://doi.org/10.1109/TVCG.2012.291

[8] N. Cao, Y. R. Lin, F. Du, and D. Wang. 2015. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications* PP, 99 (2015), 1–1. DOI: https://doi.org/10.1109/MCG.2015.73

[9] Nan Cao, Yu-Ru Lin, and David Gotz. 2016. Untangle map: Visual analysis of probabilistic multi-label data. *IEEE Transactions on Visualization and Computer Graphics* 22, 2 (2016), 1149–1163.

[10] Nan Cao, Lu Lu, Yu-Ru Lin, Fei Wang, and Zhen Wen. 2015. SocialHelix: Visual analysis of sentiment divergence in social media. *Journal of Visualization* 18, 2 (2015), 221–235. https://doi.org/10.1007/s12650-014-0246-x

[11] Nan Cao, Conglei Shi, W. Sabrina Lin, Jie Lu, Yu-Ru Lin, and Ching-Yung Lin. 2016. TargetVue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 280–289. DOI: https://doi.org/10.1109/TVCG.2015.2467196

[12] N. Cao, J. Sun, Y. R. Lin, D. Gotz, S. Liu, and H. Qu. 2010. FacetAtlas: Multifaceted visualization for rich text corpora. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov 2010), 1172–1181. DOI: https://doi.org/10.1109/TVCG.2010.154

[13] Daniel B. Carr, Richard J. Littlefield, and Wesley L. Nichloson. 1986. Scatterplot matrix techniques for large N. In *JASA*. 297–306. DOI: http://dl.acm.org/citation.cfm?id=26036.26072

[14] Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM'10)*, Washington, DC. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1538

[15] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Xiaoru Yuan, Nan Cao, and Yadong Wu. 2016. D-map: Visual analysis of ego-centric information diffusion patterns in social media. In *Proceedings of IEEE Visual Analytics Science and Technology (VAST16)*, Baltimore, MD. 41–50.

[16] Siming Chen, Lijing Lin, and Xiaoru. 2017. Social media visual analytics. *Computer Graphics Forum* 36, 3 (2017), 563–587. https://doi.org/10.1111/cgf.13211

[17] Siming Chen, Xiaoru Yuan, Zhenhuang Wang, Cong Guo, Jie Liang, Zuchao Wang, Jiawan Zhang, and Xiaolong Zhang. 2015. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 270–279. DOI: https://doi.org/10.1109/TVCG.2015.2467619

[18] Weiwei Cui, Shixia Liu, Zhuofeng Wu, and Hao Wei. 2014. How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2281–2290. DOI: https://doi.org/10.1109/TVCG.2014.2346433

[19] Tuan Nhon Dang, Nick Pendar, and Angus Forbes. 2016. TimeArcs: Visualizing fluctuations in dynamic networks. *Computer Graphics Forum* 35, 3 (2016), 61–69. https://doi.org/10.1111/cgf.12882

[20] Michele Emmer. 2005. *The Visual Mind II (Leonardo Books)*. The MIT Press.

[21] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3 (2010), 75–174.

[22] Daniel Fried and Stephen G. Kobourov. 2014. Maps of computer science. In *Proceedings of the 2014 IEEE Pacific Visualization Symposium (PacificVis'14)*, Yokohama. 113–120. DOI: https://doi.org/10.1109/PacificVis.2014.47

[23] Emden R. Gansner, Yifan Hu, and Stephen G. Kobourov. 2010. GMap: Drawing graphs as maps. In *Proceedings of the 17th International Conference on Graph Drawing (GD'09)*, Chicago. 405–407. DOI: https://doi.org/10.1007/978-3-642-11805-0_38

[24] Emden R. Gansner, Yifan Hu, and Stephen C. North. 2013. Interactive visualization of streaming text data with dynamic maps. *Journal of Graph Algorithms and Applications* 17, 4 (2013), 515–540.

[25] Adrien Guille, Hakim Hacid, Cécile Favre, and Djamel A. Zighed. 2013. Information diffusion in online social networks: A survey. *SIGMOD Record* 42, 2 (2013), 17–28. DOI: https://doi.org/10.1145/2503792.2503797

[26] Diansheng Guo, Jin Chen, Alan M. MacEachren, and Ke Liao. 2006. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (Nov 2006), 1461–1474. DOI: https://doi.org/10.1109/TVCG.2006.84

[27] Jeffrey Heer and Danah Boyd. 2005. Vizster: Visualizing online social networks. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis'05)*, Minneapolis, MN. 5. https://doi.org/10.1109/INFOVIS.2005.39

[28] Nathalie Henry and Jean-Daniel Fekete. 2007. Matlink: Enhanced matrix visualization for analyzing social networks. In *Proceedings of the IFIP Conference on Human-Computer Interaction*, Rio de Janeiro. 288–302. https://doi.org/10.1007/978-3-540-74800-7_24

[29] Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. 2007. NodeTrix: A hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1302–1309. https://doi.org/10.1109/TVCG.2007.70582

[30] Hexagonal maps. [n. d.]. http://cartonerd.blogspot.com/2015/05/helecxagon-mapping.html.

[31] Mengdie Hu, Shixia Liu, Furu Wei, Yingcai Wu, John Stasko, and Kwan-Liu Ma. 2012. Breaking news on twitter. In *CHI Conference on Human Factors in Computing Systems (CHI'12)*, Austin. 2751–2754. DOI: https://doi.org/10.1145/2207676.2208672

[32] Yifan Hu. 2005. Efficient and high quality force-directed graph drawing. *The Mathematica Journal* 10 (2005), 37–71. http://www.mathematica-journal.com/issue/v10i1/contents/graph_draw/graph_draw.pdf.

[33] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th International Workshop on Knowledge Discovery on the Web, Advances in Web Mining and Web Usage Analysis (WebKDD'07), and 1st International Workshop on Social Networks Analysis (SNA-KDD'07)*, San Jose, CA. 56–65. DOI: https://doi.org/10.1145/1348549.1348556

[34] Brian Karrer and M. E. J. Newman. 2010. Stochastic blockmodels and community structure in networks. *CoRR* abs/1008.3926 (2010). http://arxiv.org/abs/1008.3926

[35] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. 2008. Information visualization. In *Visual Analytics: Definition, Process, and Challenges*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.). Springer-Verlag, 154–175. DOI: https://doi.org/10.1007/978-3-540-70956-5_7

[36] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, and Robert C. Miller. 2011. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'11)*, Vancouver. 227–236. DOI: https://doi.org/10.1145/1978942.1978975

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

[38] Lev Nachmanson, Roman Prutkin, Bongshin Lee, Nathalie Henry Riche, Alexander E. Holroyd, and Xiaoji Chen. 2015. Graphmaps: Browsing large graphs as interactive maps. *CoRR* abs/1506.06745 (2015). http://arxiv.org/abs/1506.06745

[39] Tiago P. Peixoto. 2014. The graph-tool python library. *figshare* (2014). DOI: https://doi.org/10.6084/m9.figshare.1164194

[40] Marco Pennacchiotti and Ana-Maria Popescu. 2011. A machine learning approach to twitter user classification. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, Barcelona. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2886.

[41] Donghao Ren, Xin Zhang, Zhenhuang Wang, Jing Li, and Xiaoru Yuan. 2014. WeiboEvents: A crowd sourcing weibo visual analytic system. In *Pacific Visualization Symposium (PacificVis) Notes*, Yokohama. 330–334.

[42] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey P. Ellis, and Daniel A. Keim. 2014. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec 2014), 1604–1613. DOI: https://doi.org/10.1109/TVCG.2014.2346481

[43] Gerard Salton and Michael J. McGill. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc.

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

[45] Lei Shi, Hanghang Tong, Jie Tang, and Chuang Lin. 2014. Flow-based influence graph visual summarization. *CoRR* abs/1408.2401 (2014). http://arxiv.org/abs/1408.2401

[46] Marc A. Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, and Eric Gleave. 2009. Analyzing social media networks with NodeXL. In *Proceedings of the Fourth International Conference on Communities and Technologies, C&T*, University Park, PA. 255–264. DOI: https://doi.org/10.1145/1556460.1556497

[47] Guodao Sun, Yingcai Wu, Shixia Liu, Tai-Quan Peng, J. J. H. Zhu, and Ronghua Liang. 2014. EvoRiver: Visual analysis of topic coopetition on social media. *IEEE Transactions on Visualization and Computer Graphics.* 20, 12 (Dec 2014), 1753–1762. DOI: https://doi.org/10.1109/TVCG.2014.2346919

[48] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. 2009. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris. 807–816. DOI: https://doi.org/10.1145/1557019.1557108

[49] James J. Thomas and Kristin A. Cook (Eds.). 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* IEEE Computer Society.

[50] Ramine Tinati, Leslie Carr, Wendy Hall, and Jonny Bentwood. 2012. Identifying communicator roles in Twitter. In *Proceedings of the 21st World Wide Web Conference (WWW'12)*, Lyon. 1161–1168. DOI: https://doi.org/10.1145/2187980.2188256

[51] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605 (2008), 85.

[52] C. Vehlow, F. Beck, P. Auwärter, and D. Weiskopf. 2015. Visualizing the evolution of communities in dynamic graphs. *Computer Graphics Forum* 34, 1 (Feb. 2015), 277–288. DOI: https://doi.org/10.1111/cgf.12512

[53] Fernanda Viégas, Martin Wattenberg, Jack Hebert, Geoffrey Borggaard, Alison Cichowlas, Jonathan Feinberg, Jon Orwant, and Christopher Wren. 2013. Google+Ripples: A native visualization of information flow. In *Proceedings of the 22nd International World Wide Web Conference (WWW'13)*, Rio de Janeiro. 1389–1398. DOI: http://dl.acm.org/citation.cfm?id=2488388.2488504

[54] Xiting Wang, Shixia Liu, Yang Chen, Tai-Quan Peng, Jing Su, Jing Yang, and Baining Guo. 2016. How ideas flow across multiple social groups. In *Proceedings of IEEE Visual Analytics Science and Technology (VAST'16)*, Baltimore, MD. 51–60. https://doi.org/10.1109/VAST.2016.7883511.

[55] W. Wu, Y. Zheng, N. Cao, H. Zeng, B. Ni, H. Qu, and L. M. Ni. 2017. MobiSeg: Interactive region segmentation using heterogeneous mobility data. In *Proceedings of the 2017 IEEE Pacific Visualization Symposium (PacificVis)*. 91–100. DOI: https://doi.org/10.1109/PACIFICVIS.2017.8031583

[56] Yingcai Wu, Nan Cao, David Gotz, Yap-Peng Tan, and Daniel A. Keim. 2016. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia* 18 (2016), 2135–2148. DOI: https://doi.org/10.1109/TMM.2016.2614220

[57] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. 2014. OpinionFlow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics.* 20, 12 (2014), 1763–1772. DOI: https://doi.org/10.1109/TVCG.2014.2346920

[58] Rebecca Xiong and Judith Donath. 1999. PeopleGarden: Creating datnulla portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (UIST'99)*. 37–44. DOI: https://doi.org/10.1145/320719.322581

[59] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan J. H. Zhu, and Huamin Qu. 2013. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics.* 19, 12 (2013), 2012–2021. DOI: https://doi.org/10.1109/TVCG.2013.221

[60] Muye Yang and Robert P. Biuk-Aghai. 2015. Enhanced hexagon-tiling algorithm for map-like information visualisation. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction (VINCI'15)*. ACM, 137–142. https://doi.org/10.1145/2801040.2801056

[61] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. 2014. FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics.* 20, 12 (2014), 1773–1782. DOI: https://doi.org/10.1109/TVCG.2014.2346922

Received September 2017; revised December 2017; accepted January 2018