R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media

Shuai Chen, Sihang Li, Siming Chen, and Xiaoru Yuan, *Senior Member, IEEE*

Abstract—We propose R-Map (Reposting Map), a visual analytical approach with a map metaphor to support interactive exploration and analysis of the information reposting process in social media. A single original social media post can cause large cascades of repostings (i.e., retweets) on online networks, involving thousands, even millions of people with different opinions. Such reposting behaviors form the reposting tree, in which a node represents a message and a link represents the reposting relation. In R-Map, the reposting tree structure can be spatialized with highlighted key players and tiled nodes. The important reposting behaviors, the following relations and the semantics relations are represented as rivers, routes and bridges, respectively, in a virtual geographical space. R-Map supports a scalable overview of a large number of information repostings with semantics. Additional interactions on the map are provided to support the investigation of temporal patterns and user behaviors in the information diffusion process. We evaluate the usability and effectiveness of our system with two use cases and a formal user study.

Index Terms—Social Media, Information Diffusion, Map-like Visual Metaphor

1 Introduction

Social media has become an important platform allowing users to publish, acquire and share information. People can not only post messages, but can also reply or repost messages from others, thus accelerating the spread of information on social platforms. Once a message is posted (i.e., *the original message*), users can repost it with or without comments (i.e., *reposting messages*). Reposting messages convey a diverse variety of topics and sentiments of social media users. Key players such as opinion leaders may cause a lot of repostings, as they have many followers and are usually more influential. Information spreads via such reposting cascades, meaning that an original message can reach a broader range of social media users, who are not limited to the followers of the user posting the original message. Our work focuses on investigating reposting behaviors in social media. We focus on analyzing the diverse reposting cascades with multiple levels and different semantics.

Existing works focus on the structure [9, 30, 43, 45] or the semantics [6, 7, 48, 50] of reposting behaviors. It is challenging to design a visualization that supports an understanding of the multi-level structure and the semantics at the same time. There are three main reasons for this. Firstly, the large extend of users' participation requires a scalable visualization method to understand the structure of diffusion. Secondly, the semantics of the messages vary a great deal in the reposting structure. Different users play different roles in the reposting structure, and how to facilitate an understanding of these diverse semantics in the reposting structure is challenging. Thirdly, reposting behaviors and semantics vary over time, which makes the analysis even more complicated.

To address the above challenges, we introduce R-Map, an interactive visualization with map symbols that can reveal the reposting pattern of an original message in a spatial context (Figure 5). In this map, key players are visualized as lakes, and users reposting them become counties. Counties with similar topics form a region, while a lake and its surrounding counties form a country, which represents a subtree of the reposting tree. Connected countries form continents, and isolated islands represent messages that are different from the continents in topics. We use multiple connection symbols to represent different relations between key players. A river connecting lakes represents the reposting relations between key players. A route represents a reposting message that was not made by a follower, and a bridge represents a reposting message causing different repostings in semantics. Unlike previous map designs, we address the issue of scalability by transforming a multi-level reposting tree into a map. The map is enriched by the diverse semantic information, which is not addressed in previous work.

The key contributions in this paper include:

- **A Novel Visual Metaphor Design for Information Reposting.** The reposting map provides an overview of the reposting process. It represents multi-faceted information using intuitive visual metaphors to enable users to explore different patterns.
- **A Visual Analytics System to Explore User Reposting Behaviors in Social Media.** Different interactions are provided to help users interact with the map more easily, allowing users to understand the reposting process from different perspectives.

The rest of this paper is organized as follows. We discuss related work in Section 2 and data in Section 3. We then describe the design of R-Map and the construction method in Section 4, and give an introduction to the system introduction in Section 5. Two cases are presented to justify the usefulness of our method in Section 6. The results of a user study are reported in Section 7. Finally, we conclude with a discussion and directions for future work.

2 Related Work

The state-of-the-art works in social media visualization and visual analytics are summarized in [11, 47]. In this paper, we review related works in the fields of information diffusion, tree visualizations, and map-based visualizations.

2.1 Analysis of Information Diffusion

Online social networks play a key role in the diffusion of information. Katz and Lazarsfeld propose the "two-step flow" theory of communication [23], which asserts that information from the mass media moves to opinion leaders, and then moves to a wider range of audiences. Opinion leaders are influential in the diffusion process and pass on their interpretations in addition to the original messages.

In addition to modeling the process of information spreading [20, 21], other research studies have been conducted to predict the diffusion

- *Shuai Chen, Sihang Li, and Xiaoru Yuan are with Key Laboratory of Machine Perception (Ministry of Education), School of EECS, Peking University. E-mail: {shuai.chen, lisihang, xiaoru.yuan}@pku.edu.cn. Xiaoru Yuan is also with National Engineering Laboratory for Big Data Analysis and Application, Peking University. Xiaoru Yuan is the corresponding author.*
- *Siming Chen is with the Fraunhofer Institute IAIS and University of Bonn. E-mail: siming.chen@iais.fraunhofer.de.*

Manuscript received 31 Mar. 2019; accepted 1 Aug. 2019.
Date of publication 16 Aug. 2019; date of current version 20 Oct. 2019.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2019.2934263

of information [31, 49] and to detect diffusion patterns [38]. Unlike these works, this paper utilizes visualization techniques to illustrate the diffusion process of information using a map metaphor.

2.2 Visual Analytics for Information Diffusion

Many types of visualizations have been proposed to reveal the diffusion patterns in social media. Many of these directly visualize the reposting tree of a message with an improved node-link diagram. WeiboEvents [30] provides four different layouts (tree, circular, sail, and curve layouts) to help users analyze the diffusion structure of a weibo. Cascade [43] monitors sharing activities on Twitter using a 3D visualization. Users can smoothly change to different views to explore different facets of the information. Chen et al. analyze information diffusion with the geographic facets [10]. However, these methods may not scale to large datasets. Revisit [40] displays reposting chains using a focus+context timeline visualization, which hides reposting relations that are out of focus. Google+ Ripples [45] uses a mixture of the node-and-link and circular treemap methods. The goal of R-Map is to provide a scalable visualization method for a large scale reposting tree with multi-faceted information, which is different from previous works.

Other research studies focus on analyzing the diffusion patterns of collective messages, which often have the same topics. Whisper [6] can be used to monitor the information diffusion on a given topic in real time. Zhao et al. designed FluxFlow [50], an interactive visual analysis system to explore the anomalous spreading of information. E-Map [7] collects weibos and repostings on different stages about an event to analyze the evolution patterns. All of these methods are successful designs in terms of revealing the information diffusion process from different perspectives, but the detailed propagation patterns of a single message are neglected.

2.3 Tree Visualization

Over 300 tree visualization methods are collected at treevis.net [34]. They are divided into explicit and implicit tree visualizations according to representation methods of parent-child relations [33]. Explicit node-link diagrams using straight lines or curves are the most common tree visualization methods. Although they allow users to easily interpret the relations of nodes, node-link diagrams have a rather low space efficiency, which causes clutter when analyzing large scale reposting trees.

For implicit tree visualizations, treemaps [22] and icicle plots [25] are two prominent methods. In treemaps, hierarchical relations are represented by nesting the children wthin the parent. Users can easily compare the attributes of the data based on the sizes of nodes, but the treemap imposes high burdens on users in terms of interpreting the tree structure [3]. Although various methods have been proposed to help users gain more insights from treemaps [44, 46, 45], they are not intuitive to show the information diffusion structures. Icicle plots or sunburst [39] represent the parent/child relations through adjacency: the child nodes are placed adjacent to their parent. To ensure the adjacency of parent and children, nodes vary greatly in size. Although many tree visualizations have been proposed, they focus only on showing the structure of trees. It is hard to create an intuitive way of visualizing the features and relations of nodes in the reposting tree, which is the focus of our paper.

2.4 Map-Based Visualization for Relation Data

Maps are used to draw the surfaces of a region to represent the distribution of geographical features. Couclelis was among the earliest to investigate the potential of generating visualizations with geographical metaphors [12]. The advantages of geographical metaphors are the good user experience they offer and the solid theories of geography that underpin them. Skupin et al. [35, 37] also discussed the cartographic research on transferring metaphors to non-geographical visualizations. They mapped conference abstracts onto a term-dominance landscape [36]. The introduction of GMap [15] accelerated research in this area. GMap transforms graphs into maps to highlight communities and this method has been applied in the analysis of network communication [13], computer science literature [14] and video content [28]. In the area of social media, Gansner et al. [16] proposed a dynamic map generation method to analyze evolving topics in real-time Twitter messages. CompactMap [26] uses a spiral-based layout to generate a compact map for comparing cluster sizes of streaming tweets. Chen et al. [9] introduced D-Map, a hexagon-based map for analyzing user-centric information diffusion patterns. Later, they proposed D-Map+ [8] and E-Map [7], which allowed users to conduct a multi-faceted analysis of event evolution on social media. However, these methods provide an aggregated view of the diffusion behaviors of multiple source messages and omit the multi-level structure of repostings.

3 Social Media Data

In this section, we introduce the data and summarize the important features in the reposting process.

3.1 Data Attributes

The data in this study are taken from Sina Weibo, a social media platform like Twitter that is widely used in China. Each weibo is similar to a tweet on Twitter. We focus on analyzing the reposting process of an original weibo. The following terms are defined to describe the reposting data:

- **Root user** is the creator of the original weibo.
- **Repostings** include both direct and indirect repostings. If a weibo reposts another weibo, it is a **direct reposting**. If a weibo reposts another reposted weibo, it is an **indirect reposting** of the original weibo. The repostings of a given weibo are all weibos that directly or indirectly repost it.
- **Comments** are added by social media users to give their opinions on the reposted weibo.
- **Semantics-varying repostings** are weibos creating repostings with new topics compared to the reposted weibo.
- **Non-follower repostings** are weibos posted by users who do not follow the creator of the reposted weibo.
- **Key players** are social media users who create large-scale repostings to the original weibo.

A cascade of user repostings forms a tree. In this tree, the original weibo is the root node, and all repostings are descendants nodes of the root. Direct repostings of a node are its child nodes.

The characteristics of the dissemination of a given weibo message can be summarized using the following aspects.

- **D1: Reposting Structure.** Many users can participate in the reposting process, creating complicated tree structures.
- **D2: Key Players.** Users have different roles in the reposting process. Key players have a strong influence on shaping the reposting tree structure and semantics.
- **D3: Different Behaviors.** Repostings may have different semantics or temporal features. Here, we explore the features of reposting users as these may reveal the social mechanism underlying the process.
- **D4: Dynamic Diffusion.** The life cycle of the dissemination of a weibo has multiple stages, including its beginning, bursting, and dying. At each stage, there are different reposting behaviors, topics and key players.

3.2 Different Relations

We investigate three types of relations in our analysis of reposting behaviors: follower-followee relations, key players' reposting relations, and semantic relations.

Follower-followee relations. In social media, users' messages are normally reposted by their followers. In contrast, reposting by a non-follower may disseminate the message to a wider range of audiences, greatly changing the diffusion process. This relation (i.e., reposting but not following) is specially addressed in R-Map.

Key players' reposting relations. Reposting by key players or opinion leaders, exposes the message to the general public. Thus, we are interested in the reposting chains of key players to understand how a message can reach these key players. The reposting patterns after the key players repost a message are also analyzed in depth.

Semantic relations. Social media users may express different attitudes in their messages. These messages concern diverse topics involving different communities of social media users. The sentiments expressed in the reposting messages are also important in understanding users' emotions towards the events.

4 Visualization Design

In this section, we first discuss the motivation and design rationales for R-Map. We then present the visual encoding methods and steps in map construction.

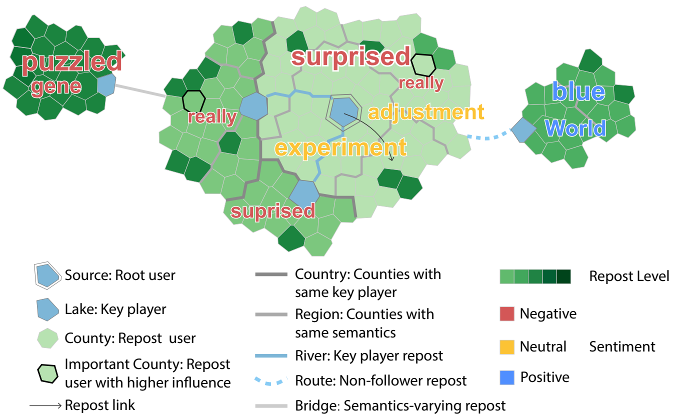

Fig. 1. Visual encodings of R-Map, which is constructed based on a reposting tree.

4.1 Motivation: Semantic Map

Existing visualization methods fail to provide an intuitive representation of the reposting tree and semantics, so a new visual design is needed to address the challenges of visualizing a reposting tree. Maps describe the physical living spaces of people. Since their emergence, many different types of maps have been created to show different kinds of information. Experts have confirmed that people generally learn to read maps in pre-school [4], and need little training in solving map-based visualization tasks. In our work, we use maps to organize information derived from the reposting tree, and link them with spatial objects such as continents, islands, countries, rivers, etc. The convenience of maps in terms of cognition means that people can better understand reposting data from multiple perspectives. Moreover, the spatialization of information in the map provides opportunities to visualize both the structure and the semantics at the same time.

4.2 Design Rationale

Our analysis of reposting trees focuses on the reposting structure, the roles of users, and the evolution of the semantics. We summarize the following requirements to present these key factors and their relations.

- **R1: Providing an overview of the reposting structure.** The map should show the overall structure of the reposting tree to help people understand the diffusion process of the original weibo (D1). The map should be easy to understand and should avoid visual clutter.
- **R2: Identifying users with different roles and behaviors.** Users play different roles and have different attitudes in the diffusion process. Key players should be differentiated from ordinary users (D2). We need to reveal users' sentiments and opinions of their messages in the diffusion process (D3).
- **R3: Revealing the dynamic diffusion process.** The topics and the sentiments of users may vary at different stages. The messages of key players can create new opinions or topics, and may have a strong influence on the overall event (D4).
- **R4: Coupling the analysis of reposting structure and semantics.** The semantics of repostings affect the development of the reposting tree. These should be considered together in the design in order to reduce the analysis burden on users when switching between different contexts.

4.3 Map Design and Visual Encoding

R-Map provides an overview of the structure and the semantics for the diffusion of an original weibo. Reposting features are mapped to different metaphors on the map, including lakes, counties, regions, countries, continents, islands, rivers, routes, and bridges (Figure 1). We choose these metaphors because they are common in a real map and can be linked in a meaningful way to the data features.

Lake: In the map, a lake represents a key player in the diffusion process. A lake is the origin of rivers, and thus irrigates the surrounding land, in the same way as a key player disseminates information to users who repost him/her weibos. The root user is placed in the center of the country, and is distinguished from other nodes with an additional stroke. For other key players, they are placed on the boundary to show reposting relations (R1, R2).

County: A county on the map is a polygon representing a non-keyplayer weibo node in the reposting tree. A map can be divided into different counties, just as a reposting tree consists of different nodes. Counties with higher influence are highlighted with a black stroke. A county is affiliated with a lake and belongs to a country (R1).

Region: Counties discussing similar topics form regions. The repostings of a weibo may contain discussions of different topics by social media users. We group these repostings according to topics. Counties with the same topics form continuous regions on the map (R2, R4).

Country: A country on the map symbolizes a subtree rooted by a key player. All counties associated with the country are reposting nodes on the subtree. The position of a country is determined by a tree layout algorithm [18] and reveals the overall diffusion structure (R1). A country may have multiple regions.

Continent and Island: Large connected countries on the map form a continent. Smaller and isolated countries form islands. The connections between different countries represent different reposting relations (R1).

River, Route and Bridge: These features encode different reposting relations of the key players discussed in Section 3.2. **Rivers** connect different lakes on the continent, and represent the parent-child relations of key players, thus showing the skeleton of the reposting tree. Reposting often happens in a follower network [41] and indicates an agreement with the message content [29]. However, there are always non-follower repostings and semantics-varying repostings in the diffusion process. To highlight these special repostings for key players, we split the territories of key players from the continents in the form of islands. If a key player does not follow his/her parent, a curved **route** connects them; if the topics of the territories belonging to the key player are different from his/her parent, a straight **bridge** connects them. Bridges and routes are common in a map and can be easily distinguished. Both a bridge and a route can connect to an island at the same time (R3, R4). Only the different relations among key players are highlighted, as too many links will make the map degrade into a node-link tree with much visual clutter. Users can interactively check the reposting relations of other nodes in the system.

In the proposed map design, we omit most of the links of the original tree. To compensate for the loss of structural information (i.e., the level of the nodes in the tree), we use color to encode the level information. The reposting levels are mapped to colors with the same hue but with monotonically increasing saturation values [42]. A green-scale color scheme is employed to make the map more intuitive and comprehensible. The darker the color of the polygon, the deeper it is within the reposting tree structure (R1). An alternative color scheme is also provided to represent the sentiment of each reposting: red, orange, and blue are used to represent negative, neutral and positive opinions, respectively (Figure 9). Repostings without comments are considered to be neutral. The sentiment of each reposting is calculated using an on-

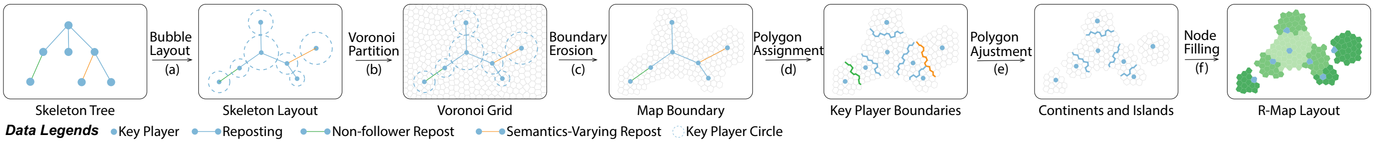

Fig. 2. R-Map construction. (a) The position of each key player node is calculated using a bubble tree layout method [18]. (b) The layout bounding box is partitioned by a Voronoi Diagram. (c) The generated polygons are then filtered based on the density value given in Equation 3. (d) Each polygon is then assigned to a key player node based on the weighted distance function in Equation 4. (e) To get the final layout area, the polygons assigned to each key player are adjusted further. (f) Finally, all nodes are filled into the calculated areas.

line Chinese sentiment analysis tool [1] (R3). The color coding methods of map metaphors are based on guidance provided by the United States Department of the Interior [49] and the color scheme of Google Maps. Rivers, bridges, and routes are color-coded with blue, gray, and blue, respectively. To denote the boundaries of different countries, we use thick dark lines, and for regional boundaries, we use thinner gray lines to distinguish them from the country boundaries. The white space of the canvas is color-coded with light blue, similar to the color of the ocean.

To enable a better understanding of the semantics in the diffusion process, keywords are overlapped on the map with color indicating different sentiments (R4).

4.4 R-Map Construction

The construction of R-Map first determines certain key nodes in the tree and then places the remaining nodes, in a similar way to existing graph algorithms such as HDE [19] and Pivot MDS [5]. They are determined by a threshold value based on the number of the descendant nodes (repostings). We set the default threshold T as 5% of the total number of the nodes after testing. Using this setting, if one node is a key node, its parent will also be considered a key node. We can extract a connected subtree from the original reposting tree. The subtree is considered to be the skeleton of the original tree. Similarly, counties with higher influence within each country are nodes with repostings more than 1% of the total number of the nodes. Users can change the threshold interactively in our system. The construction of R-Map includes two steps: skeleton layout and map layout (Figure 2).

4.4.1 Skeleton Layout

In this step, the skeleton layout determines the basic shape of the generated map. The following layout requirements are considered. Firstly, the layout should show the tree structure clearly, and the path from the root to a node should avoid a zigzag shape. Secondly, each key node should be allocated a certain area, so that its descendant nodes can be filled in. We adopt the bubble tree drawing algorithm proposed by Grivet et al. [18]. The radius (r_i) of the circle for the key node (K_i) is calculated as follows:

$$r_i := \sqrt{K_i.descendants.length}, \tag{1}$$

where $K_i.descendants$ contains the descendant nodes of K_i, which will be placed in the circular area allocated to K_i. The descendant nodes of K_i do not include the descendant nodes of K_j if K_j is a child node of K_i. In the original algorithm, circles with a parent-child relation touch externally. However, in our method, to ensure the generation of islands, an extra distance is added between two adjacent circles with non-follower or semantics-varying relation (Figure 2c). The extra distance value d is set as half of r_m (the minimal radius value of circles) after tuning. For two key players with a parent-child relation, if the child does not follow the parent, the boundary distance between them is set as d (the follower-followee relation). We calculate the keyword vector for each key player based on its messages and those of its descendants. If the cosine distance between the parent and child is less than 0.5, this indicates they have a different topic distribution. The boundary distance between them is also set as d (semantic relation).

4.4.2 Map Layout

After obtaining the positions of the key players, we calculate the positions of the remaining nodes in the reposting trees. The input is the layout of the skeleton tree, and the output is the position of each node in the original reposting tree. Each node is represented as a polygon on the map.

Plane Partition: (Figure 2b) Based on the layout of the skeleton tree, the bounding box of the drawing canvas is first calculated. Following this, the plane is divided by a Voronoi Diagram. The Voronoi Diagram generates many polygons, which are used to encode nodes on the reposting tree. The Voronoi Diagram was chosen because it is irregular and similar to a real-world terrain. We randomly project the seed points into the bounding box and apply the Lloyd algorithm [27] to generate a relatively uniform size of polygons. The number of seed points (N_s) is calculated as:

$$N_s := W * H/\pi, \tag{2}$$

in which W and H are the width and height of the bounding box size, respectively. We set the area of each polygon as π in Equation 2, so that the number of polygons within a key player circle is almost equal to the number of its descendant nodes.

Boundary Erosion: (Figure 2c) Although the generated polygons completely occupy the plane, not all of them will be used. Based on the positions of the key players, we carry out splatting of them into the polygons using a Gaussian Kernel. The density value of each polygon in the Voronoi Diagram is calculated as:

$$D := \sum_{i=1}^{N_k} \alpha_n \mathcal{N}(P|\mu_i, {\sigma_i}^2), \tag{3}$$

where N_k is the number of key players. α is the weight of each key player, which corresponds to the number of descendants. μ is the position of each key player and σ is the radius of the Gaussian Kernel. The splatting radius σ is equal to the radius of the key player circle. The density values of all polygons are then normalized. We set the threshold value to 0.1 after tuning. All polygons with a density below the threshold are discarded. In this step, we obtain the basic shape of the map.

Polygon Assignment: (Figure 2d) After creating the shape of the map, each polygon is assigned to a key player node. We allocate polygons based on a weighted distance function, in a similar way to the Voronoi treemap construction method [2]. For each polygon, for which the center position is denoted as p, its distance to a key player node with coordinate $K.pos$ is:

$$distance(p, r, K.pos) := ||p - K.pos|| - r, \tag{4}$$

where r is the radius of the key player circle, which is used as a weight in the distance function. Each polygon is assigned to its nearest key player node. The map is divided into different continuous areas after this assignment.

Polygon Adjustment: (Figure 2e) We have allocated a different number of polygons to each key player. However, the number of allocated polygons is not strictly equal to that of the nodes to be placed. The polygons and nodes of the reposting tree have a one-to-one mapping. The extra polygons for some key players will leave blank areas inside the map, while some key players may have insufficient polygons to place their repostings. We apply a two-round iteration method to adjust the allocation of polygons so that each key player node has a number of polygons equal to the number of descendant nodes. In the first round, if the number of polygons allocated to a key player is insufficient, its territories are expanded by adding polygons along the boundary. Once the allocated polygons are enough, their affiliations will not change. In the second round, key players with extra polygons have these polygons iteratively removed. In this step, we first identify key players with territories that are land-locked by others. The extra polygons are assigned to their neighboring key players. For other key players, the polygons are removed along the boundary, according to their distance from the center of the territory. The polygons furthest away will be removed first. After the two rounds of adjustment, the whole map is divided into continents and islands as designed.

Node Filling: (Figure 2f) After adjusting the polygons, the lake position representing each key player is first determined. The root node is assigned the nearest polygon to its location. Other key player nodes are assigned the polygon on the boundary which is nearest to the position of its parent polygon. The key player polygon on the boundary serves as a port to the country in which it is located. The descendant nodes of each key player are then placed within each country.

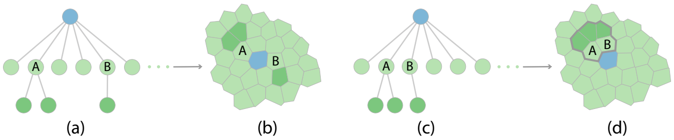

Fig. 3. Chronological layout(a, b) and semantic layout (c, d). (a) The original reposting tree. (b) Nodes are placed chronologically around the parent node. (c) Subtree B with similar semantics to subtree A, is adjusted in the layout sequence. (d) Subtrees A and B form a region on the map.

For the descendant nodes of each key player, their layout should maintain the reposting relations. Firstly, nodes with the same ancestor are placed within a continuous area. Secondly, nodes should be close to their parents on the plane. In the Voronoi Diagram, a polygon has a limited number of adjacent polygons, meaning that for a node with many children, not all of the child polygons will be adjacent to the parent polygon; they are placed around the parent node, instead. Two layout strategies are considered. In a **chronological layout,** We first sort nodes on the same level chronologically. Then, when each node has been placed on the map, its child nodes are placed successively. This layout uses a depth-first strategy, so that a subtree can take up a continuous area on the map, as shown in Figure 3b. In this method, the semantics of each node are not considered and no regions appear in the generated map. In the **semantic layout**, in order to reveal the group behaviors of nodes, nodes with similar keywords are placed together in a region on the map. For each subtree rooted by a child node of the key player, we extract keywords from the messages of nodes and build a weighted vector using the Term Frequency-Inverse Document Frequency (TF-IDF) [32] method. We then apply k-means clustering [27] to group subtrees into different clusters. The cluster number k is determined by the elbow method [24]. Subtrees without comments in reposting are grouped and placed first, so that users can compare the activeness of commenting behavior for different key players. Other clusters are placed chronologically based on the average reposting time. In Figure 3, subtrees rooted by A and B are placed together and form a region on the map (Figure 3d). Compared with the chronological layout, the semantic layout highlights the diverse distribution of topics in the messages reposting key players.

4.4.3 Boundaries and Links

After generating the map layout, map metaphors, including boundaries, rivers, routes and bridges, are created to highlight the reposting structures on the map.

Boundary: Edges owned by two polygons of different countries form the boundaries of countries. Edges owned by two polygons of different regions form the boundaries of regions.

River: Rivers connect two key players (lakes). Each river travels along the grid points of the polygons and it is the shortest path between the source and target key players.

Route and Bridge: Routes and bridges connect different lands. A child key player positioned at the coast is connected to the nearest polygon on the coast of the parent key player territory.

Figure 4 shows a comparison of the different thresholds that generate different numbers of countries. Lower thresholds generate more countries, which show more details of the reposting structure but take up more space.

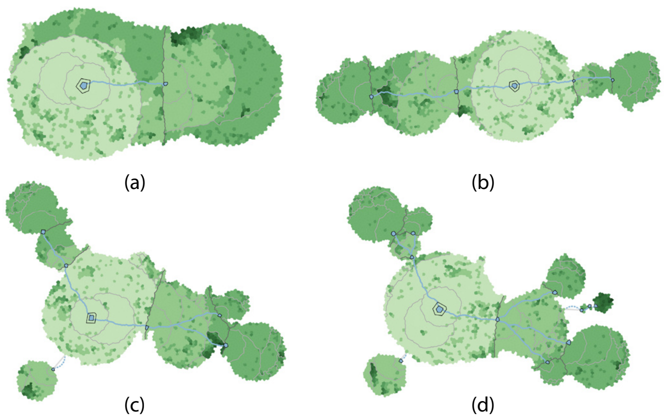

Fig. 4. Different thresholds for extracting key players. The original weibo has 2,563 repostings in total. With a decrease in the threshold, more countries are generated. (a) T = 1000, generating two countries; (b) T = 300, generating five countries; (c) T = 100, generating seven countries; (d) T is 30, generating 12 countries.

5 System Interface

The visual analytics system includes a Map View, a Weibo Table View, a Word Cloud View, and a Timeline View to support the visual exploration of the reposting diffusion process. (Figure 5).

5.1 Map View

The Map View (Figure 5b) provides an overview of the reposting structure and the different topics discussed in the reposting process (R1). Rivers, bridges and routes explicitly indicate the different relations between key players. Each country is further divided into regions representing different topics. Repostings without comments are located around the key players. Important counties are highlighted in each region, and are accompanied by some darker areas, which are their descendant repostings. Users can click the county on the map to investigate the diffusion path and features of the reposter, as shown in the User Information Panel (Figure 5c). These features include the delay time (F1) from the original weibo, the relation with the parent node (F2), the numbers of direct repostings (F3) and total repostings (F4), and the numbers of followers (F5) and followees (F6). F1 and F2 help users to compare the behaviors of different reposters, while F3 and F4 indicate the influence of reposters. F5 and F6 help users to understand the reason for the popularity of messages. A prior study by Suh et al. [41] proved that the numbers of followers and followees, URLs and hashtags have strong relations with reposting. URLs and hashtags are shown in the text of the message. The border color of the panel indicates the sentiment of this message. Keywords are directly overlapped on each region. This reduces the burden on users to switch between different contexts (R4). A map legend is provided to help users read the map (Figure 5d). Users can control the display of the visual elements according to their needs by toggling the radio button. Users can also explore the map by navigation, panning and zooming in the same way as with real-word online maps.

5.2 Weibo Table View

The Table View (Figure 5a) provides a list of all weibos. Users can sort weibos according to reposting time, number of repostings, or level on the reposting tree (R2). Clicking one message on the table shows the information panel for that reposter on the map. Users can also search weibos using keywords.

5.3 Word Cloud View

The Word Cloud View shows the keyword distribution of the selected weibos. The sentiment of each word is averaged by the sentiment of weibos containing the word. The size of each word denotes the

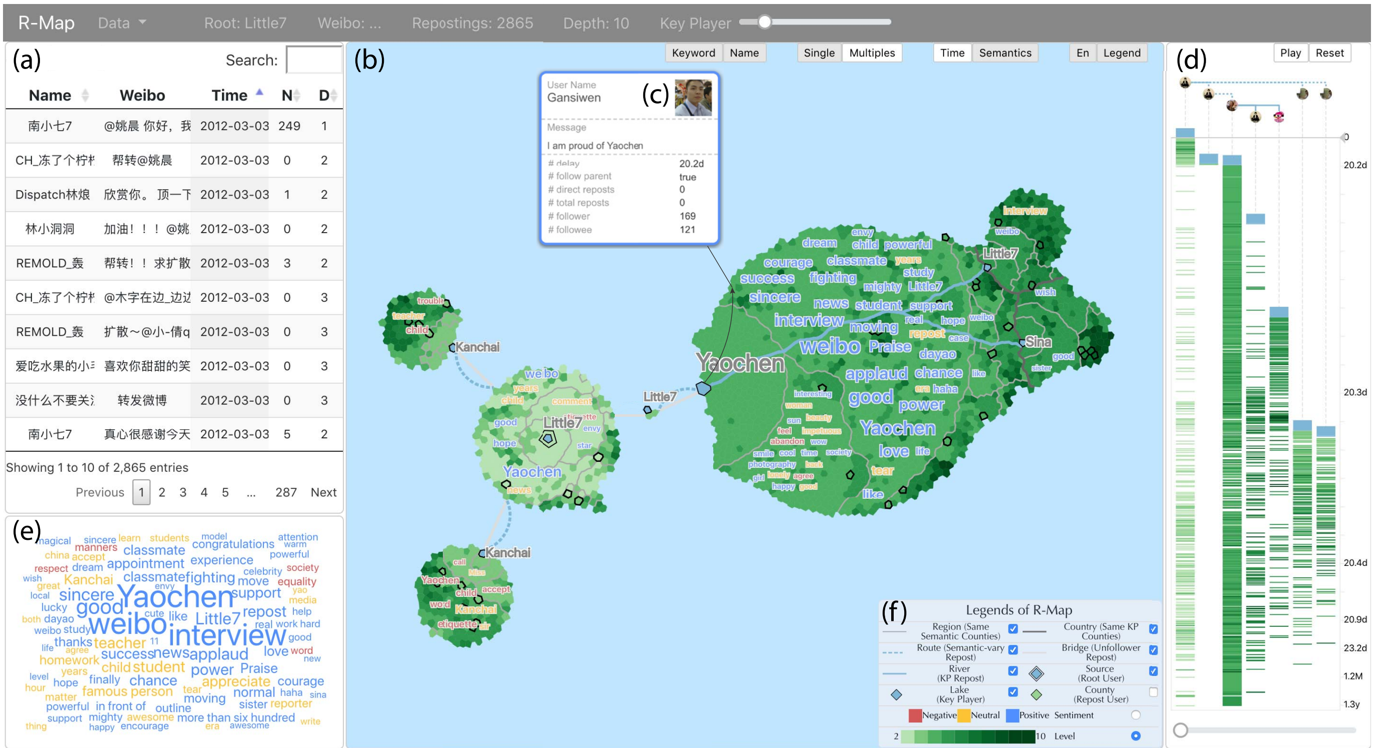

Fig. 5. R-Map System Interface. (a) Weibo Table View, for selecting different weibos; (b) R-Map View, summarizing the diffusion process of an original weibo; (c) User Information Panel, showing features of users; (d) Timeline View, revealing the temporal patterns of repostings induced by different key players; (e) Word Cloud View, showing hot keywords of the selected weibos; (f) R-Map Legend, showing visual mappings in the map. For the convenience of non-Chinese readers, keywords and some important weibos are translated from Chinese to English.

frequency it appears in the selected weibos. Users can select weibos that contain a specific word by clicking the word in this view. The Word Cloud View is dynamically updated to show keywords of new messages when an animation of the diffusion process is played (R3).

5.4 Timeline View

The Timeline View (Figure 5c) provides a temporal overview of the diffusion process (R3). Above the timeline, a node-link diagram is provided to indicate the reposting relations of the key players, which are sorted according to the reposting time. Below each key player (blue rectangle), all their descendant messages are listed based on the reposting time. The time ticks show the delay time of reposting from that of the original weibo. An animation function is provided to fast-forward the diffusion process. A brush function is also provided for exploring repostings of interest.

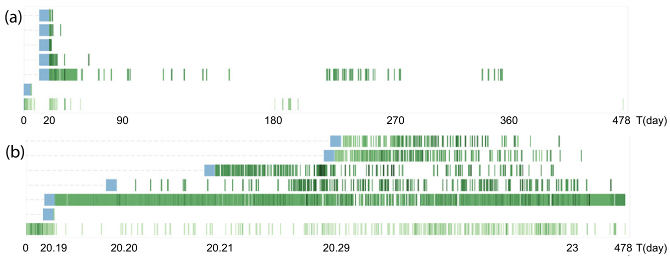

Fig. 6. Timeline equalization: comparison between (a) the original timeline and (b) the equalized one. With equalization, the detailed temporal patterns are shown more clearly.

Repostings on social media often have a long time diffusion process, which results in uneven temporal distribution. Mapping the repostings linearly on the plane according to the reposting time will result in dense overlapping. In this work, we adopt histogram equalization, which is widely used for contrast enhancement in image processing [17], to equalize the temporal mapping of repostings. The total time of the reposting process is divided into equal-length periods according to the sample time interval. Repostings that fall into the same period form a time block. We assign a continuous interval on the timeline to each time block using histogram equalization. Inside each block, each reposting is mapped linearly based on its reposting time. The original timeline and the equalized timeline are compared in Figure 6, which shows that the latter has much less overlapping.

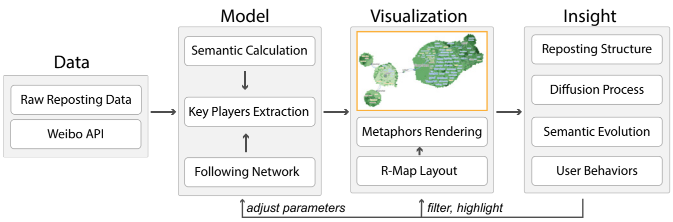

Fig. 7. Workflow of the R-Map visualization system.

All these views are highly coordinated to support the exploration of weibo repostings based on the system workflow shown in (Figure 7).

6 Usage Scenarios

We present two use cases based on real data from Sina Weibo and demonstrate how our system can reveal the reposting process.

6.1 Case 1: Influence of Key Players

In the first case, we use R-Map to illustrate how key players can enlarge the influence of the source message and promote the evolution of topics.

Map Overview. We can gain an overview of the event from the generated map shown in Figure 5b. The original weibo was posted by *Little7* (the source lake), and was reposted by *Kanchai* (twice) and again by the original poster. *Little7* was then reposted by *Yaochen*, creating a large area of territories on the map. *Yaochen* is a famous Chinese actress and has millions of followers on Sina Weibo. *Little7*

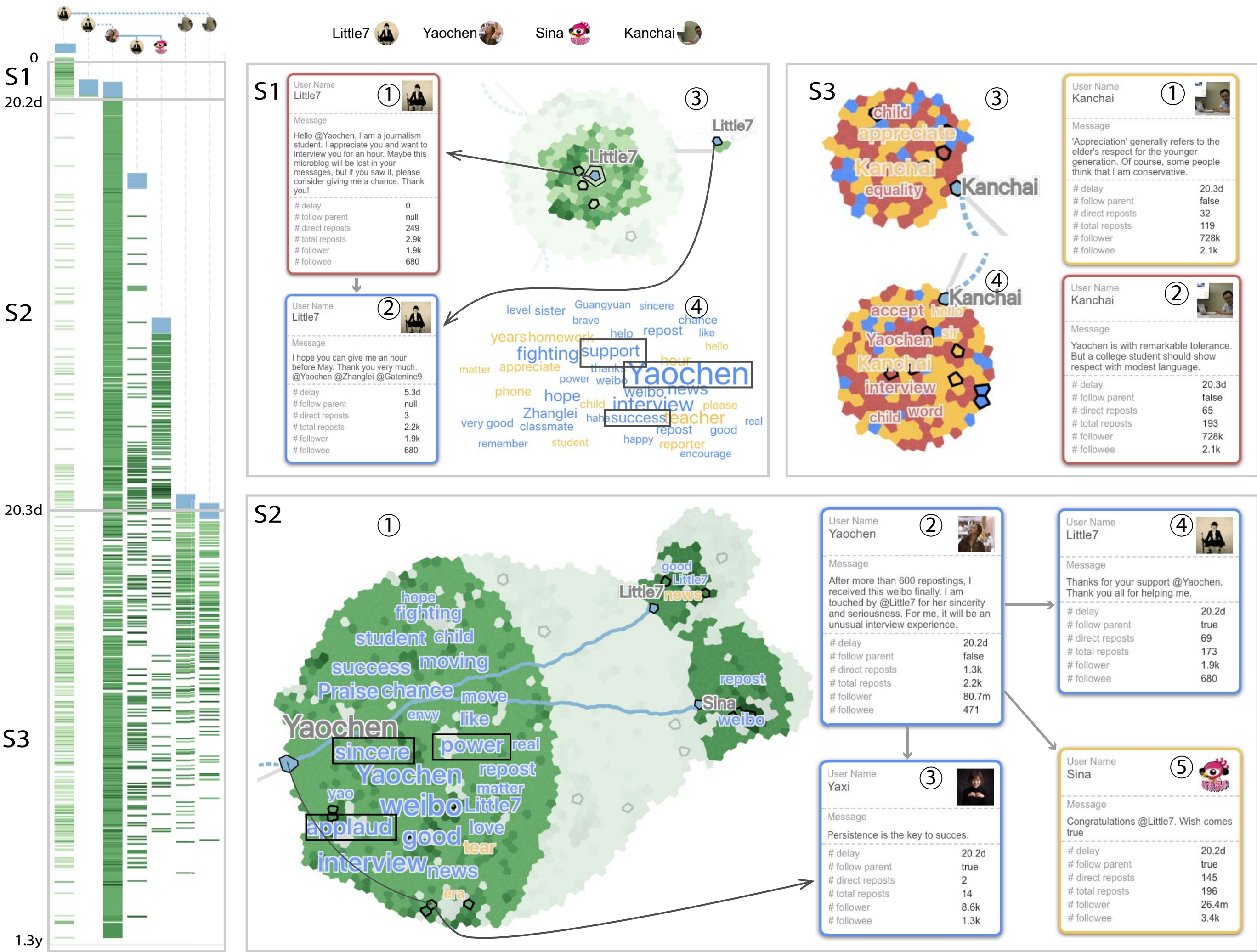

Fig. 8. The case of *Little7*, visualizing the influence of the key players in the reposting process. By exploring the map view in Figure 5, we identified three stages in the reposting process, including the initial stage (S1), the bursting stage (S2) and the final stage with diverse range of discussions (S3).

and *Sina* also reposted *Yaochen*'s weibo. The key players reposting *Little7* did not follow her, and introduced different discussion topics, as indicated by the routes and bridges. The map was generated using the semantic layout. Countries are divided into different regions and different keywords are shown. Most of the keywords are positive, and negative keywords are mostly shown on the territories associated with *Kanchai*. After fast-forwarding the diffusion process using the Timeline View, we divided the event into three stages (Figure 8 - S1, S2, and S3).

Dynamic Exploration of the Event Stage (S1) - An Initial Scenario. By brushing the Timeline View, we found that at the first stage (Figure 8-S1), *Little7* posted the original message expressing her wish to interview the famous actress *Yaochen* (S1-1). Although she tagged *Yaochen* with the "@" function, she did not expect that this message would be seen by *Yaochen*, who receives thousands of weibos every day. A small number of people reposted the message, constructing the terrain around the root node on the map (S1-3). Most of these encouraged her and tagged *Yaochen* in their repostings, as indicated in the Word Cloud View (S1-4). *Little7* then reposted her first weibo, generating a second lake on the map. Clicking this lake shows that this weibo explained her proposal for the interview in more detail, and tagged *Yaochen* again (S1-2). However, few people responded to this weibo.

Dynamic Exploration of the Event Stage (S2) - Involvement of the Key Player. The repostings then almost stopped for a period of time. Almost 20 days later, things changed (Figure 8-S2). *Yaochen* reposted the message from *Little7*, as shown on the map, which created numerous repostings, generating a big country. We examined *Yaochen*'s information (S2-2) and found that she had over 80 million followers on Sina Weibo. In contrast, she only followed 471 users. *Yaochen* did not follow *Little7*, so they are connected by a route on the map. *Yaochen*'s response to *Little7* created hot discussions. Keywords such as "applaud", "power" and "sincere" appear in these repostings (S2-1), which are different from those in *Little7*'s repostings, as the bridge between *Yaochen* and *Little7* explicitly indicates. Inside the country, some important counties could be found. For example, a user called *Yaxi* was amazed at the story, saying "Persistence is the key to success" (S2-3). Another fourteen users reposted her. *Little7* responded and expressed her gratitude to *Yaochen* within a short time (S2-4). A key player called *Sina* also reposted this news to its audience (S2-5). *Little7* and *Sina* were followers of *Yaochen*, and the keywords used in their message's repostings were similar. Countries affiliated with them form a continent on the map. The discussions were almost all positive at this stage. Social media users praised *Yaochen* for her kindness and congratulated *Little7* on her success. The whole stage of S2 lasted for only a few hours, but initiated numerous discussions. Compared with S1, S2 demonstrates the great influence of key players on the diffusion of information.

Dynamic Exploration of the Event Stage (S3) - Diverse Semantic Exploration. At the last stage (Figure 8-S3), another user called *Kanchai* participated in the discussions. He reposted the original weibo twice within a short time (S3-1, S3-2), generating two different

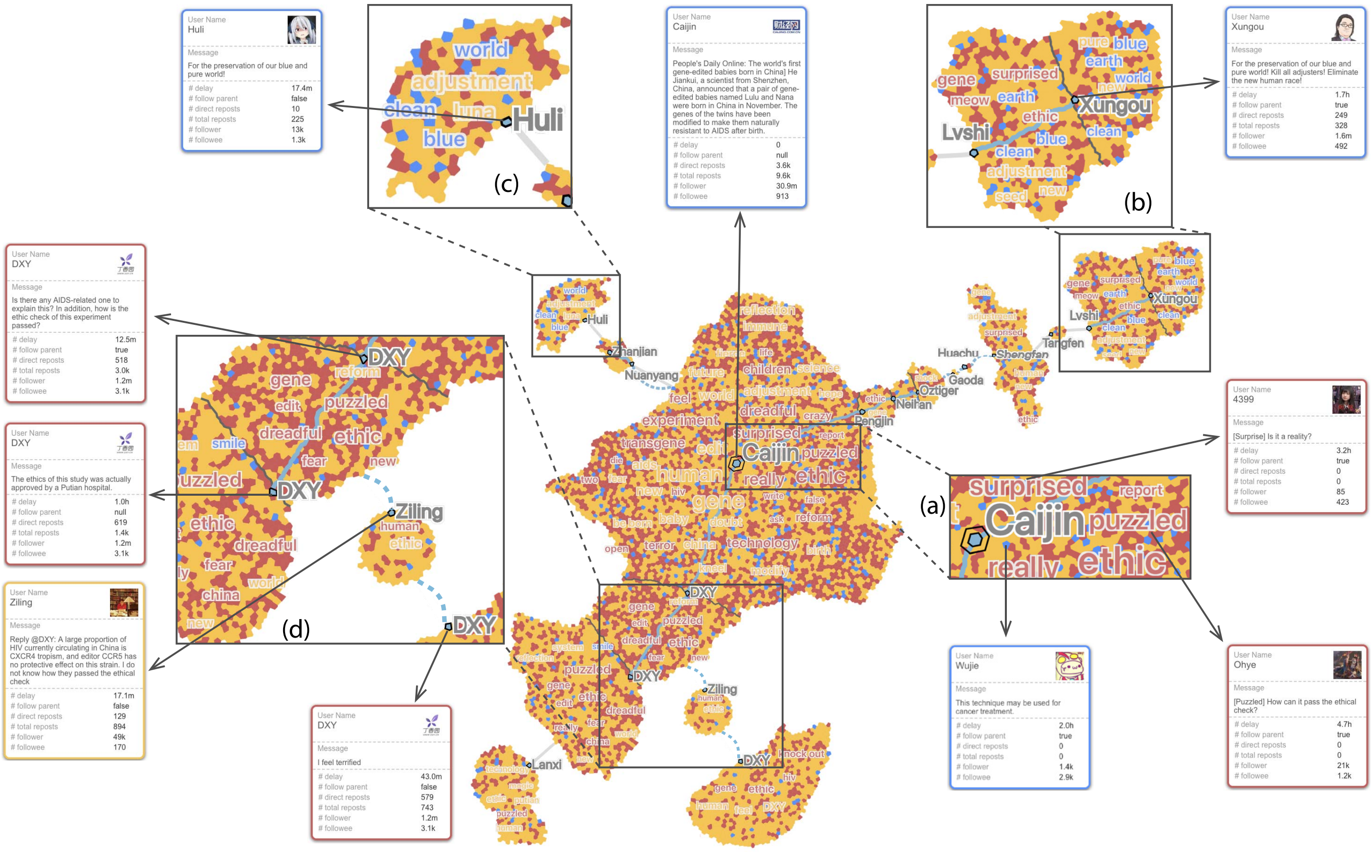

Fig. 9. The case of semantic interpretation of weibos for gene-edited babies. Most of the repostings expressed negative or neutral attitudes. (a) The original message from *Caijin*, announcing that the world's first gene-edited babies were born; (b, c) A special groups of people reposting with irony. (d) Discussion about ethic issues.

islands on the map (S3-3, S3-4). Both of these islands are connected to *Little7* by a route and a bridge, indicating that *Kanchai* did not follow *Little7* and introduced new topics to the original weibo. We checked his weibos, and found that he criticized *Little7* for showing a lack of respect to *Yaochen*. In one island, he pointed out that *Little7* did not use modest language in the original message (S3-2); in the other, he stated that the word "appreciate" was not suitable for a young girl to use to an older woman (S3-1). He had over 70,000 followers, and his weibos raised arguments about equality and politeness in interpersonal relationships. By coloring the map with sentiments, we found that negative sentiments dominated counties in these islands.

In this case, by analyzing a message from an ordinary person, we can show how the original message diffused to a large number of people through reposting by key players, who not only change the temporal patterns of reposting but also introduce different sentiments and topics. Compared with traditional tree visualization methods, R-Map can intuitively reveal semantic features in the reposting process while showing the diffusion tree structure.

6.2 Case 2: Sentiment and Semantics Interpretation

In the second case, we analyze the reposting process of a hot weibo. The map, which is color-coded by sentiment, gives us a **sentiment overview** of the reposting process and the opinions of weibo users (Figure 9). There are few positive (blue) counties on the map, and most people expressed negative (red) opinions. Three main branches extend out from the center of the map. By investigating the source lake, we found that the root user *Caijin* with nearly 40 million followers, posted the original weibo that the world's first gene-edited babies were born in China. This message initiated heated discussions and over 9,000 people reposted this message within a short time.

Zoom in and Explore Detailed Topics on Demand - Continent. We zoomed into the map and found that keywords related to the event such as "human", "ethic", and "gene" had a high frequency. People also showed surprise and concern in relation to this experiment (Figure 9a). After examining these keywords, related weibos were identified. One person posted "[Surprise] Is it a reality?" ([Surprise] indicates an emotion icon expressing surprise). Another person felt puzzled, saying "How can it pass the ethical check?". In contrast, few repostings had positive sentiments. One user reposted with the comment "This technique may be used for cancer treatment".

Zoom in and Explore Detailed Topics on Demand - Islands. We navigated to two distant islands from the continent (Figure 9b, c). From the word clouds on both islands, we found that these two islands shared many common keywords, such as "world", "blue" and "clean", which are all rare positive words. Some users' comments are very similar, containing the phrase "For the preservation of our blue and pure world". After further study, we found that this sentence was taken from a comic called "GUNDAM SEED", which tells the story of a war between humans and coordinators, who are genetically modified humans. When people used this sentence in their comments, they were expressing concern and irony about this event.

Zoom in and Explore Detailed Topics on Demand - Behaviors of Key Players. Further valuable information was found in the lower-left corner of the continent (Figure 9d). Following the rivers, we found a key player named *DXY* (a well-known medical group), with 1.2 million followers. It reposted the original weibo and questioned the ethical checks on this experiment. In the messages reposting it, a user called *Ziling*, who did not follow *DXY*, analyzed this event from a professional perspective, and stated that the experiment did not pass ethical checks. *DXY* then reposted his weibo and reported this fact to its audience. Although they were not followers of each other (connected by a route), they had a heated discussion on this topic, which involved many people. Later, *DXY* reposted its previous weibo and added more information about this study. The statement that "The ethics of this study were approved by a Putian hospital" caused further discussions with negative sentiments, since Putian has a bad reputation in China.

In this case, by analyzing weibos about a hot event, we can see that although the topics of repostings varied, they were related to the central event. Repostings relating to similar topics are gathered together on the map. Additionally, user behaviors such as discussions with influential people can make the reposting structure more complex.

7 Evaluation

To evaluate R-Map, we conducted a user study to examine how well users could understand the metaphors of R-Map and exploit the system to analyze the reposting process of an original weibo. We invited 14 participants (nine male and five female) with basic visualization knowledge. After taking part in a 10-minute tutorial, the participants were required to analyze social media data from the case in Section 6.1. We recorded participants' answers and completion time for each task, followed by a questionnaire with seven questions regarding the rationales of metaphor design and the usability of the system.

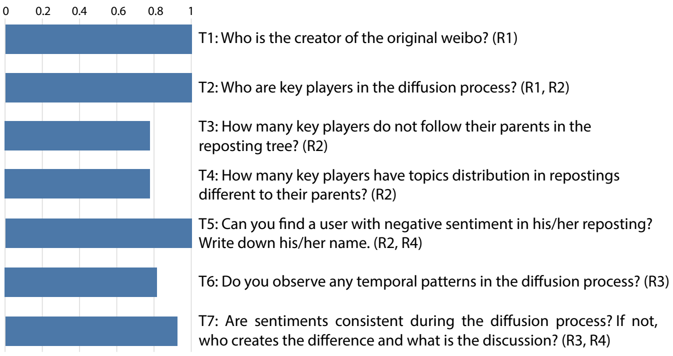

Fig. 10. Evaluation tasks and passing rates.

Figure 10 shows the seven tasks and the corresponding level of accuracy. Participants finished all the tasks within 15 minutes on average. The accuracies for T1, T2, and T5 were all 100%, demonstrating the effectiveness of R-Map in revealing the different behaviors of users. For T3 and T4, some users misread the meaning of rivers and bridges, which led to wrong answers. For T6, some participants did not identify the three main diffusion stages because they neglected the nonlinear mapping of the Timeline View. For T7, some participants ignored the reposting time of each key player. They thought *Yaochen* reposted *Little7* after the reposting by *Kanchai*, so the weibo of *Yaochen* changed sentiment from negative to positive.

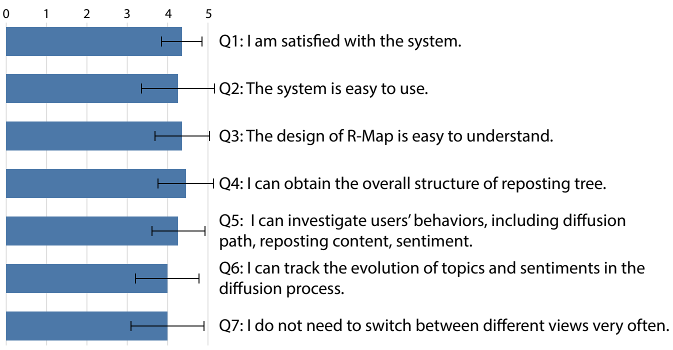

Fig. 11. Analysis of the ratings. All questions received an average rating of no less than four, which was very encouraging.

Figure 11 shows the scores for seven aspects evaluated by participants. In general, participants were satisfied with the system and could understand the metaphors easily on the map. However, the ratings for Q6 and Q7 were lower than for the others. This may be because the temporal information was not easy to track on the 2D layout compared to a 1D timeline.

Participants were interviewed about the pros and cons of the system. Most participants felt the map design was intuitive and creative. They also provided reasonable suggestions for improving the system, giving feedback such as "I think the calculation of sentiment could be improved." "I can not obtain the detailed information about the diffusion path effectively" These form good start points for further work.

8 Discussion

In the field of visual analytics, semantic map construction is attracting more and more research focus. We address the differences in analysis tasks, visual metaphor, and novelty between R-Map and several recently proposed semantic map series, such as D-Map [9], D-Map+ [8], and E-Map [7]. D-Map focuses on ego-centric information diffusion, while D-Map+ and E-Map focus on event analysis. In contrast, R-Map focuses on in-depth diffusion analysis of a single weibo. In terms of visual metaphor, D-Map and D-Map+ use the hexagon map design, while E-Map and R-Map use a natural map form. Compared with E-Map, countries, regions, and counties have different meanings in R-Map. R-Map also uses two unique metaphors, bridges and routes. The construction of R-Map takes into consideration both the reposting structure and the semantics, while previous works do not consider them simultaneously.

R-Map provides intuitive map metaphors to help in the exploration of reposting data from the perspective of both structure and semantics. Although novel and powerful, the proposed visualization has some limitations. Firstly, the shape of the generated map relies on the extracted key players; when the extracted key players change, the shape of the map will also change. More key players can provide more insight from the overview of the map, but the layout will not be compact, and require more space. We therefore need to find a balance between space efficiency and structure. To compensate for the loss of structural information, a chronological layout is proposed that uses color encoding the level of nodes. We also highlight some important nodes inside each country to guide users to explore the reposting structure of the tree. Secondly, too many visual properties on the map may introduce visual clutter. The metaphors and word cloud used in the map provide sufficient information, but users may feel overwhelmed. We allow users to control the display of the visual elements on the map according to their needs. Thirdly, the sentiment detection method could be improved using the state-of-the-art natural language processing work.

R-Map is designed to analyze Weibo reposting data, but it can also be applied to other social media data, such as Twitter. A tweet can be retweeted by Twiter users, which is similar to reposting on Weibo. We also envision the further development of R-Map as a general approach that is not limited to social media data. For example, R-Map could be used to visualize a company structure, in which the source lake is the founder of the company, and lakes are heads of different departments. Each department (country) could be divided into different groups (regions), and rivers, routes, and bridges could encode business relationships between different departments. We intend to apply our system to these areas and to evaluate our method in conjunction with domain experts.

9 Conclusion

We propose a novel visualization method called R-Map that provides an intuitive summary of the information reposting process. We transform the reposting features into different map metaphors and construct a vivid map. The diffusion structure can be visualized using various link metaphors, including rivers, routes and bridges. The distribution of topics and sentiments are also visualized on the map using different regions and colors. We develop a visual analytics system to reveal the diffusion patterns of real-world dataset. We present two cases and a user study to demonstrate the effectiveness and usability of R-Map. In future work, we will extend R-Map to other data and domains.

Acknowledgments

The authors thank the anonymous reviewers for their valuable comments. This work is supported by the National Key Research and Development Program of China (2016QY02D0304) and the National Program on Key Basic Research Project (973 Program) No. 2015CB352503, as well as PKU-Qihoo Joint Data Visual Analytics Research Center.

REFERENCES

[1] Chinese sentiment analysis tool provided by baidu open platform. http://ai.baidu.com/tech/nlp/sentiment_classify.

[2] M. Balzer and O. Deussen. Voronoi treemaps. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, pp. 49–56, 2005. doi: 10.1109/INFVIS.2005.1532128

[3] S. T. Barlow and P. Neville. A comparison of 2-d visualizations of hierarchies. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, pp. 131–138, 2001. doi: 10.1109/INFVIS.2001.963290

[4] M. Blades, J. M. Blaut, Z. Darvizeh, S. Elguea, S. Sowden, D. Soni, C. Spencer, D. Stea, R. Surajpaul, and D. Uttal. A cross-cultural study of young children's mapping abilities. *Transactions of the Institute of British Geographers*, 23(2):269–277, 1998.

[5] U. Brandes and C. Pich. Eigensolver methods for progressive multidimensional scaling of large data. In *International Symposium on Graph Drawing*, pp. 42–53. Springer, 2006. doi: 10.1007/978-3-540-70904-6

[6] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2649–2658, 2012. doi: 10.1109/TVCG.2012.291

[7] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, and X. Zhang. E-map: A visual analytics approach for exploring significant event evolutions in social media. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pp. 36–47, 2017. doi: 10.1109/VAST.2017.8585638

[8] S. Chen, S. Chen, Z. Wang, J. Liang, Y. Wu, and X. Yuan. D-map+: Interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. *ACM TIST*, 10(1):11:1–11:26, 2019. doi: 10.1145/3183347

[9] 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 *2016 IEEE Conference on Visual Analytics Science and Technology, VAST 2016, Baltimore, MD, USA, October 23-28, 2016*, pp. 41–50, 2016. doi: 10.1109/VAST.2016.7883510

[10] S. Chen, J. Li, G. Andrienko, N. Andrienko, Y. Wang, P. H. Nguyen, and C. Turkay. Supporting story synthesis: Bridging the gap between visual analytics and storytelling. *IEEE transactions on visualization and computer graphics*, 2019. To Appear.

[11] S. Chen, L. Lin, and X. Yuan. Social media visual analytics. *Computer Graphics Forum*, 36(3):563–587, 2017. doi: 10.1111/cgf.13211

[12] H. Couclelis. Worlds of information: The geographic metaphor in the visualization of complex information. *Cartography and geographic information systems*, 25(4):209–220, 1998.

[13] J. J. Fowler, T. Johnson, P. Simonetto, M. Schneider, C. Acedo, S. G. Kobourov, and L. Lazos. Imap: visualizing network activity over internet maps. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, pp. 80–87, 2014. doi: 10.1145/2671491.2671501

[14] D. Fried and S. G. Kobourov. Maps of computer science. *CoRR*, abs/1304.2681, 2013.

[15] E. R. Gansner, Y. Hu, and S. G. Kobourov. Gmap: Visualizing graphs and clusters as maps. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 201–208, 2010. doi: 10.1109/PACIFICVIS.2010.5429590

[16] E. R. Gansner, Y. Hu, and S. North. Visualizing streaming text data with dynamic graphs and maps. In *Proceedings of Graph Drawing*, pp. 439–450, 2012.

[17] R. C. González and R. E. Woods. *Digital image processing, 3rd Edition*. Pearson Education, 2008.

[18] S. Grivet, D. Auber, J. Domenger, and G. Melançon. Bubble tree drawing algorithm. In *Proceedings of International Conference on Computer Vision and Graphics*, pp. 633–641, 2004. doi: 10.1007/1-4020-4179-9

[19] D. Harel and Y. Koren. Graph drawing by high-dimensional embedding. In *Proceedings of International Symposium on Graph Drawing*, pp. 207–219, 2002. doi: 10.1007/3-540-36151-0

[20] C. Ho, C. Li, and S. Lin. Modeling and visualizing information propagation in a micro-blogging platform. In *Proceedings of International Conference on Advances in Social Networks Analysis and Mining*, pp. 328–335, 2011. doi: 10.1109/ASONAM.2011.37

[21] Y. Hu, R. J. Song, and M. Chen. Modeling for information diffusion in online social networks via hydrodynamics. *IEEE Access*, 5:128–135, 2017. doi: 10.1109/ACCESS.2016.2605009

[22] B. Johnson and B. Shneiderman. Tree maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Conference on Visualization*, pp. 284–291, 1991. doi: 10.1109/VISUAL.1991.175815

[23] E. Katz, P. F. Lazarsfeld, and E. Roper. *Personal influence: The part played by people in the flow of mass communications*. Routledge, 2017.

[24] D. J. Ketchen and C. L. Shook. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*, 17(6):441–458, 1996.

[25] J. B. Kruskal and J. M. Landwehr. Icicle plots: Better displays for hierarchical clustering. *The American Statistician*, 37(2):162–168, 1983.

[26] X. Liu, Y. Hu, S. C. North, and H. Shen. Compactmap: A mental map preserving visual interface for streaming text data. In *Proceedings of IEEE International Conference on Big Data*, pp. 48–55, 2013. doi: 10.1109/BigData.2013.6691714

[27] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–136, 1982. doi: 10.1109/TIT.1982.1056489

[28] C. Ma, Y. Liu, G. Zhao, and H. Wang. Visualizing and analyzing video content with interactive scalable maps. *IEEE Transactions on Multimedia*, 18(11):2171–2183, 2016. doi: 10.1109/TMM.2016.2614229

[29] P. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn. What do retweets indicate? results from user survey and meta-review of research. In *Proceedings of Ninth International AAAI Conference on Web and Social Media*, pp. 658–661, 2015.

[30] D. Ren, X. Zhang, Z. Wang, J. Li, and X. Yuan. Weiboevents: A crowd sourcing weibo visual analytic system. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 330–334, 2014. doi: 10.1109/PacificVis.2014.38

[31] K. Saito, R. Nakano, and M. Kimura. Prediction of information diffusion probabilities for independent cascade model. In *Knowledge-Based Intelligent Information and Engineering Systems, 12th International Conference, KES 2008, Zagreb, Croatia, September 3-5, 2008, Proceedings, Part III*, pp. 67–75, 2008. doi: 10.1007/978-3-540-85567-5

[32] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, 1984.

[33] H. Schulz, S. Hadlak, and H. Schumann. The design space of implicit hierarchy visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):393–411, 2011. doi: 10.1109/TVCG.2010.79

[34] H. J. Schulz. Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications*, 31(6):11–15, 2011. doi: 10.1109/MCG.2011.103

[35] A. Skupin. From metaphor to method: Cartographic perspectives on information visualization. In *Proceedings of IEEE Symposium on Information Visualization 2000*, pp. 91–97, 2000. doi: 10.1109/INFVIS.2000.885095

[36] A. Skupin. A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications*, 22(1):50–58, 2002.

[37] A. Skupin and S. I. Fabrikant. Spatialization methods: a cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science*, 30(2):99–119, 2003.

[38] Y. Song, Z. Wen, C. Lin, and R. Davis. One-class conditional random fields for sequential anomaly detection. In *Proceedings of Joint Conference on Artificial Intelligence*, pp. 1685–1691, 2013.

[39] J. T. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Proceedings of IEEE Symposium on Information Visualization*, pp. 57–65, 2000. doi: 10.1109/INFVIS.2000.885091

[40] M. Stefaner. Revist. http://truth-and-beauty.net/projects/revisit/, [Online; accessed 2019-03-31].

[41] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *Proceedings of IEEE International Conference on Social Computing*, pp. 177–184, 2010. doi: 10.1109/SocialCom.2010.33

[42] K. M. Thyng, C. A. Greene, R. D. Hetland, H. M. Zimmerle, and S. F. DiMarco. True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, 29(3):9–13, 2016.

[43] T. N. Y. Times. Cascade. http://nytlabs.com/projects/cascade.html, [Online; accessed 2019-03-31].

[44] J. J. van Wijk and H. van de Wetering. Cushion treemaps: Visualization of hierarchical information. In *Proceedings of IEEE Symposium on Information Visualization*, pp. 73–78, 1999. doi: 10.1109/INFVIS.1999.801860

[45] F. B. Viégas, M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant, and C. R. Wren. Google+ripples: a native visualization of information flow. In *Proceedings of International Confer-*

ence on World Wide Web, pp. 1389–1398, 2013. doi: 10.1145/2488388.2488504

[46] W. Wang, H. H. Wang, G. Dai, and H. Wang. Visualization of large hierarchical data by circle packing. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 517–520, 2006. doi: 10.1145/1124772.1124851

[47] Y. Wu, N. Cao, D. Gotz, Y. Tan, and D. A. Keim. A survey on visual analytics of social media data. *IEEE Transactions on Multimedia*, 18(11):2135–2148, 2016. doi: 10.1109/TMM.2016.2614220

[48] 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

[49] J. Zhang, J. Tang, J. Li, Y. Liu, and C. Xing. Who influenced you? predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data*, 9(3):25:1–25:26, 2015. doi: 10.1145/2700398

[50] 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