DOI: 10.1111/cgf.13211

EuroVis 2017
M. Meyer, S. Takahashi, A. Vilanova
(Guest Editors)

Volume 36 (2017), Number 3
STAR – State of The Art Report

Social Media Visual Analytics

Siming Chen[1], Lijing Lin[1] and Xiaoru Yuan[†1]

[1]Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University, China

Abstract

With the development of social media (e.g. Twitter, Flickr, Foursquare, Sina Weibo, etc.), a large number of people are now using them and post microblogs, messages and multi-media information. The everyday usage of social media results in big open social media data. The data offer fruitful information and reflect social behaviors of people. There is much visualization and visual analytics research on such data. We collect state-of-the-art research and put it into three main categories: social network, spatial temporal information and text analysis. We further summarize the visual analytics pipeline for the social media, combining the above categories and supporting complex tasks. With these techniques, social media analytics can apply to multiple disciplines. We summarize the applications and public tools to further investigate the challenges and trends.

1. Introduction

Social media are web-based platforms where users create and share messages via virtual communities and social networks. In recent years, the social media change the way people communicate, share, live, etc. The core parts of social media are users and their behaviors. Users can post and repost (i.e. resend the messages initially posted by others) messages, which have time stamps, text, media, and possibly geo-tags. These behaviors lead to information diffusion in social media. The user-generated content spreads through social communication online. Social media data records all the messages posted and behaviors of users. These data are quite big and with many unseen patterns inside. Moreover, a large amount of open social media data is available. Thus, many researchers pay attention to social media analytics. Data mining can effectively identify specifically defined features on social media [TSWY09, PP11], such as influencer identification, user classification. However, not all patterns are well defined and the analysis requires a large involvement of humans. Thus, researches in visual analytics propose many advanced methods and tools to seek patterns on social media and solve problems in the analyzing process. Our work aims to summarize the state-of-the-art in visualization and visual analytics, to give a research outline and to discuss possible directions and challenges of future research in social media visual analytics.

1.1. Related Surveys

There are several surveys on analyzing and mining the behaviors in information diffusion [GHFZ13, BT14]. However, to the best of our knowledge, there are only two general reviews for social media visual analytics from Schreck et al. [SK13] and Wu et al. [WCG*16]. In 2013, Schreck et al. described a small number of representative papers in detail. We believe that a more complete survey of state-of-the-art work is necessary. Wu et al. summarized more papers from two research domains of multimedia and visualization [WCG*16]. They emphasized gathering information and analyzing user behaviors in multimedia analysis. However, we have a different perspective for collecting related works and propose a new taxonomy for classifying the visualization and visual analytics process of social media.

We also take a broader view of related surveys into consideration. These surveys include multi-variate visualization [MGMZ14], dynamic network visualization [BBDW16], text visualization [WSJ*14, KK15, KKRS13], community detection and visualization [VBW15] and personal visualization [HTA*15] etc. They have only mentioned some works in social media visual analytics and refer to them as examples in the application areas. To fill in the blank and provide an overview of related research, it is necessary to provide such a state-of-the-art survey focusing on visual analytics of social media data.

1.2. Data

There is a variety of derived data based on users' activities in social media. We investigate multiple attributes of data and propose our categories of the targeting entities.

In one aspect, users follow other users based on the existing relationship, similar hobbies, and information feed, etc., which constructs the users' following-followee network. We abbreviate it to the follower network. Users' communication and reposting behav-

† {siming.chen, lijing.lin, xiaoru.yuan}@pku.edu.cn

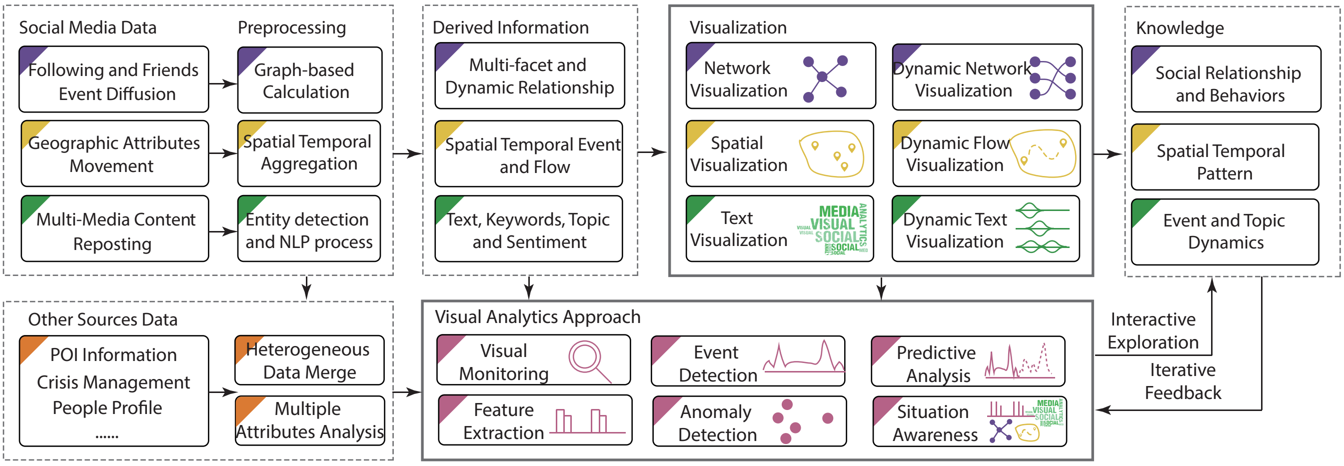

Figure 1: *Taxonomy of this survey, addressing the visualization, visual analytics techniques, applications and systems. We discuss the social media data characteristics. We derive multi-facet and dynamic networks, spatial temporal event and flow as well as text related information. We collect the research works from these three perspectives. Combined with multiple visualization techniques, we summarize the visual analytics goals and categorize them into six types, including visual monitoring, feature extraction, event detection, anomaly detection, predictive analytics, situation awareness. Visual analytics systems fulfilling these goals have applications in multiple disciplines to enable users gain knowledge.*

iors enable information diffusion and build up the reposting network [WHMW11] and diffusion network [VWH*13]. A node in reposting network represents a social media user while a node in the diffusion network is a social media message . In the mean time, user-generated content provides the semantics of users' behaviors. Such semantics can be reflected in multiple levels, including keywords [AGCH11] and topics [DWS*12]. Users also represent their sentiments by posting positive or negative messages [ZGWZ14].

The other important aspect of social media is the spatial temporal information. There are two types of such information in social media, including the rough living places as attributes indicated by users, and geo-tagged messages with precise GPS information. On one hand, we can know how messages are discussed and diffuse across different regions, cities or even countries, by examining where the participating users come from [CLS*12,ZLW13]. On the other hand, a geo-tagged social media message is a message with location information. The distribution of semantic messages with geo-tags enables users to understand the social events' spatial temporal distribution [BTH*13, MRJ*11]. By connecting geo-tagged messages per individual in a chronological order, we can roughly construct his/her trajectories with uncertainties [CYW*16]. Considering the above main aspects of social media data, we categorize the targeting data in social media into three types of entities, including network, geographic information, and text. The first type includes users' social network and information diffusion network. The users' social network is constructed by the following behaviors and reposting behaviors. The second type includes the spatial temporal information diffusion and events distribution, as well as the movement constructed from the geo-tagged messages. The last type includes keywords, topics and sentiments, which are derived from content in the social media (Figure 2).

1.3. Taxonomy of the Survey

In this survey, we contribute a taxonomy of visual analytics in social media. The overall structure of the analytical pipeline is summarized (Figure 1). We discuss visualization and visual analytics techniques, as well as the application domains in detail.

- **Entity Taxonomy in Social Media Visualization** We extract three main types of entities, including network, spatial temporal information and text (Figure 2). Each entity includes three subcategories. We discuss the corresponding visualization techniques for each entity. This part is discussed in Section 2.
- **Social Media Visual Analytics Taxonomy** We review how research works combine multiple visualization and interaction techniques to solve problems in social media visual analytics. We extract six general goals, including visual monitoring, features extraction, event detection, anomaly detection, predictive analysis and situation awareness (Figure 11). This part is discussed in Section 3.
- **Domain-specific Applications and Representative Systems** Social media visual analytics shed insights in multiple disciplines. We summarize multiple disciplines including social science theory and application, journalism, disaster management, crisis and emergency management, politics, finance, sports and entertainment, and tourism and urban planning (Figure 13). Besides, we discuss public and commercial systems (Figure 15). This part is discussed in Section 4.

2. Visualization Techniques

In this section, we discuss research works based on the categorized entities. For each type of an entity, we survey the related visualization techniques (Figure 7). Research papers may focus on one or multiple entities. We select related papers discussing social media

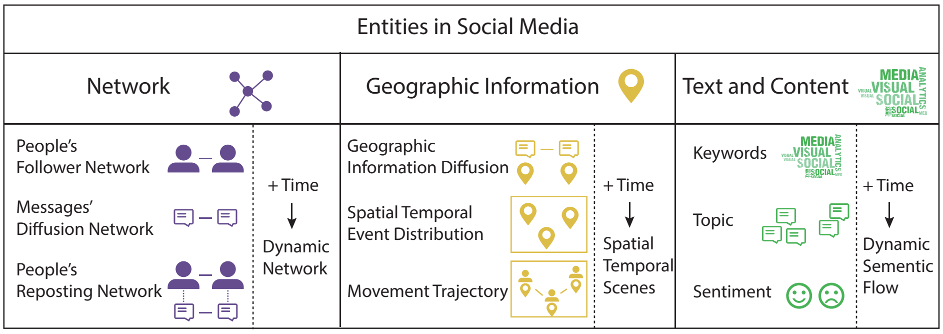

Figure 2: *Taxonomy of entities in social media. It includes three categories with three sub-categories each. As network entities, users' social network includes their follower network and reposting network. The diffusion process of messages is recorded in the diffusion network. For the geographic information, analysts need to identify where information is reposted, in order to analyze the spatial temporal information diffusion, detect events distribution and analyze people's trajectories. Keywords, topics and sentiment are important features derived from messages in social media. Moreover, dynamic features are important to derive insight for each entity.*

visual analytics and mark each paper with a main category based on its core contribution. The color in its title indicates its main category (Figure 3). If one paper deals with multiple data explicitly and applies related techniques to achieve results, we give it multiple color marks in the related categories. We detail the problems, challenges and proposed visualizations for each entity in the following sections.

2.1. Network Visualization

On social media platforms, users contact each other through following each others, reposting and commenting messages. There are two basic kinds of entities in the network of social media, namely, users and messages (i.e. tweets on Twitter). These entities are linked based on various relationships. For users, on the one hand, they follow different users based on their social relationships, interest, etc., which builds up the follower network. On the other hand, they post messages and repost others' messages, which constructs the reposting network. For messages, users' reposting behaviors lead to information diffusion, which constructs the diffusion network of messages. With the combination of different entities and various relationships, the network is complex and attracts the attention of researchers.

We define a general form, $G = (V, E)$, for networks in social media, where $V = \{v_i, v_2, \ldots, v_n\}$ is a set of entities, such as users or messages, and $E = \{e_1, e_2, \ldots, e_m\}$ is their relationship, including following, reposting and mentioning. As these relationships are all directed, we define $e_k = \{v_i, v_j\}$ as the edge starting from the node v_i and ending to the node v_j. By default, the networks we discuss in this chapter are directed networks.

2.1.1. Follower Network

In the follower network $G = (V, E)$, the node $v_i \in V$ represents a user and the edge $e_k = (v_i, v_j) \in E$ means the user v_j follows the user v_i. This network describes the social relationship between users in social media. In the last decade, lots of researchers have focused on such networks to explore social structures, community relations, etc. By extending what has been done in social network analysis in social media, we briefly introduce related network visualization methods as background knowledge.

In the emergence of social media, Heer and Boyd proposed Vizster [HB05], a system for users to explore large-scale online social networks and communities (Figure 4a). However, since the size of social networks grows largely, both the readability and the scalability of the layout become issues. Particularly, the node-link diagram has more overlapping nodes and crossing links when the number of nodes and links is increasing. In order to improve the readability and the scalability, Shen et al. [SMER06] present a visual analytical tool, OntoVis. The tool allows users to conduct structural and semantic abstraction to simplify large networks, analyze the backbone of the network and facilitate analytic reasoning for users' relationships. In 2007, Microsoft Excel spreadsheet software added NodeXL [SSMF*09], a toolkit for network overview, discovery and exploration. Besides node-link diagrams, researchers also use matrices to visualize social networks in social media. The matrix diagram, in which rows and columns are nodes and each cell represents the corresponding edge, has a high space utilization [BBDW14]. It can be used to visualize a dense graph without visual clutter [GFC05, KEC06]. Henry and Fekete propose MatrixExplorer [HF06], a system offering both the node-link diagram and the matrix representation to help users explore social networks. However, it is not easy to explore the network structure by directly juxtaposing node-link and matrix visualization. To conquer this problem, they subsequently propose MatLink [HF07], a hybrid representation with links overlaid on the top and left of a matrix. Their system works well for path-related tasks, such as finding common neighbors, shortest paths, the largest clique in the social networks. Afterwards, Henry et al. [HFM07] merge these two representations in a view, called NodeTrix. NodeTrix combines the advantages of node-link and matrix diagrams. The node-link diagram is used to visualize the global structure of the network, while a matrix-based layout can show structures in each community. The connections in each community are relatively dense. Using NodeTrix, users can explore communities structures and relationships in a more convenient way.

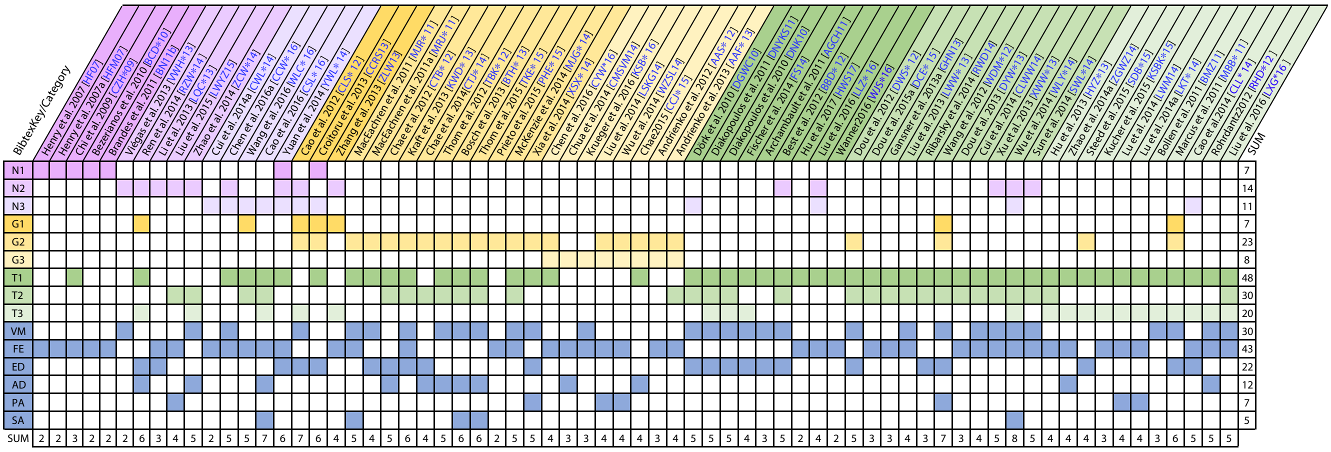

Figure 3: *68 selected papers of visualization and visual analytics for social media. N1-N3: follower network, diffusion network and reposting network; G1-G3: geographic information diffusion, spatial temporal event distribution, and movement trajectory; T1-T3: keywords, topic, and sentiment. The lower six rows are different visual analytics goals. VM: visual monitoring; FE: feature extraction; ED: event detection; AD: anomaly detection; PA: predictive analysis; SA: situation awareness. It shows the multiple entities and the corresponding visualization techniques used in each work. These works fit in one or multiple visual analytics goals.*

The network based on the follower relationship among users is a multivariate social network. Both nodes and links have additional attributes, such as name and location of users in the node attributes, as well as how long two people follow each other in the link attributes, etc. Chi et al. [CZH*09] propose a framework, iOLAP, for analyzing the multi-variate network data in the social media. They identify four main variables, including people, relation, content and time. However, this system does not support users to compare the difference and similarity between various dimensions. Bezerianos et al. [BCD*10] present GraphDict, a system for exploring the multivariate social network with a matrix, which allows users to compare different attributes easily, such as age, gender and location. Apart from being a communication platform, social media is also a suitable platform to collect demographic information. Dou et al. [DCE*15] collect demographic information, including the participant's gender, gender expression, age group, education, current location, income level, religious affiliation, etc., based on user-generated content. They present DemographicVis, a visual analytic system to support interactive analysis of demographic groups with defined features.

As relationships between users change constantly, the follower network is a dynamic graph in a long time scale. Brandes et al. [BN11] propose gestaltmatrix, which uses a glyph matrix to visualize the evolution of the relationship between each pair of users (Figure 4b). They design a gestalt-based glyph which is a metaphor of a seesaw to encode the relational data and superimpose glyphs from top to bottom in a chronological order. Users can analyze how stable the relationships are in social media.

2.1.2. Diffusion Network

In the diffusion network $G = (V,E)$, the node $v_i \in V$ represents a message and the edge $e_k = (v_i, v_j) \in E$ means the message v_j mentioning or referring the message v_i. It shows how information spreads. Related research can be generally categorized into two kinds, investigating the layout and the semantics of the diffusion network.

Starting from a source message, reposting nodes build a multi-level hierarchical structure. Hierarchy is a special relational structure, which can be visualized as node-link diagram. Google+Ripples [VWH*13] visualizes the information flow based on a hybrid of a node-link and a circular map metaphor (Figure 4c). Users can easily highlight the important messages by identifying its size and diffusion paths. Besides the circular packing, different layouts of such diffusion network shows the propagation patterns of the network from different perspectives. WeiboEvents [RZW*14] provides three layouts, including a tree layout, a circular layout, and a sail layout, which supports users to explore the information diffusion process. The circular layout suits for highlighting the overall diffusion patterns and key players. The tree layout emphasizes the hierarchical features such as depth of the hierarchical structure while the sail layouts highlight the time order (Figure 4d). Similar to the sail layout, Li et al. [LQC*13] visualize reposting relationships and time order with parallel coordinates to illustrate the evolution of the events efficiently. To further investigate the features of the diffusion process, Liu et al. [LWYZ15] regard a social media as an artificial physical system and apply a dynamic fluid model. Their method can detect the speed and the scale of information diffusion.

With the development of social media and convenience of posting, reposting and commenting messages, more and more messages arise and spread quickly every day. However, the advertising robots and rumors are also increasing. Zhao et al. [ZCW*14] extract anomalies from a huge crowd of messages based on advanced machine learning algorithms (one-class conditional random fields, OCCRF). They also propose FluxFlow, an interactive system to reveal and analyze anomalous information spreading on social media. By combining the semantic information, users could judge the de-

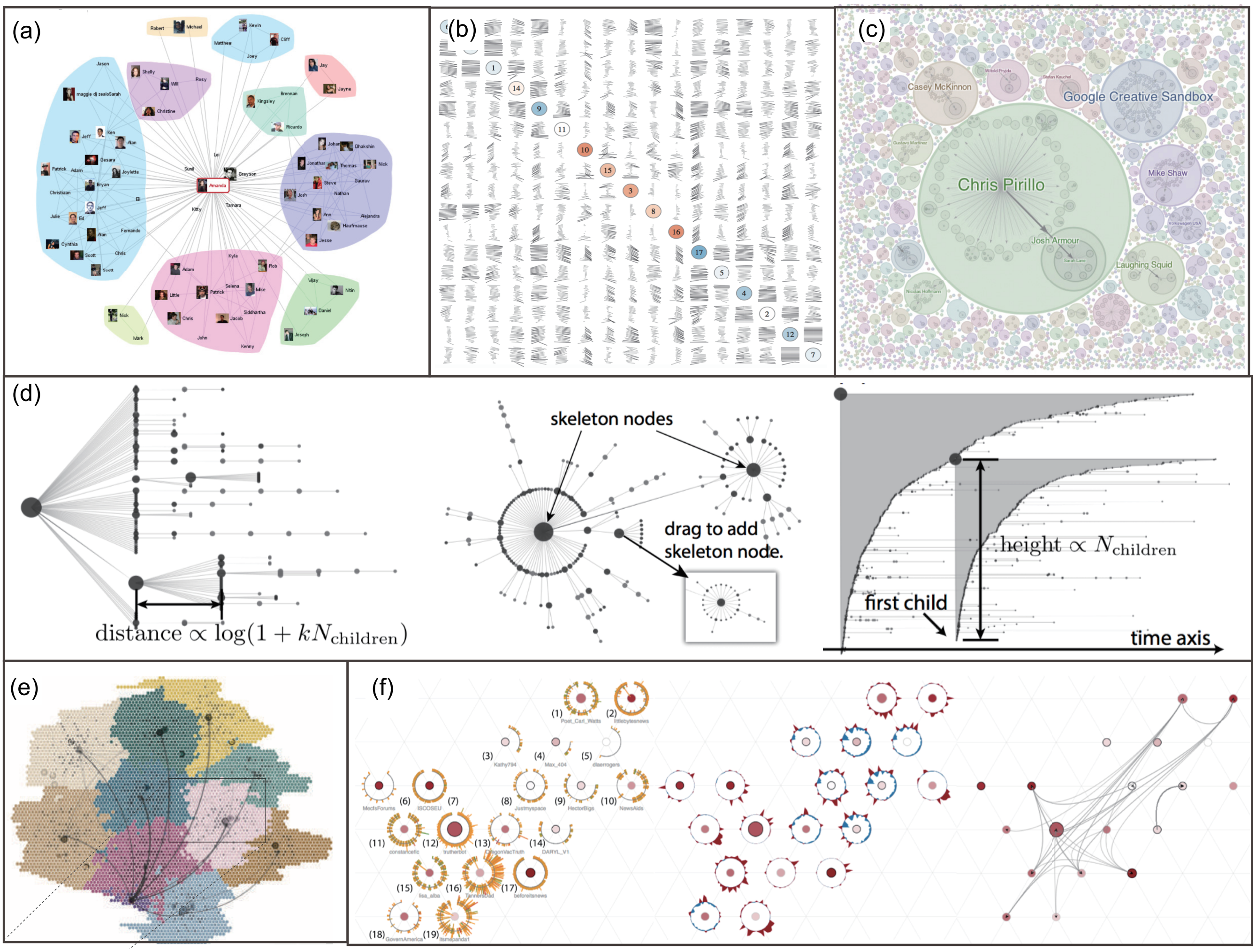

Figure 4: *Visualizations of social networks in social media. (1) Follower network visualization: Vister [HB05], showing relationships with a node-link diagram (a). Gestaltmatrix [BN11], a matrix-based layout visualizing the evolution of relationships between every two users (b). (2) Information diffusion network visualization: Google+ Ripples [VWH*13], visualizing the diffusion among users with circular packing techniques (c). WeiboEvents [RZW*14], combined with a tree layout, a circular layout and a sail layout to show multiple aspects of the information diffusion process (d). (3) Reposting network visualization: D-Map [CCW*16], a map-like visualization showing the ego-centric information diffusion among social communities (e). TargetVue [CSL*16], with circle-based glyphs to identify communication activities of anomalous users (f).*

gree of abnormality of the messages. Thus, analysts are able to take actions to stop the diffusion of anomalous information in time.

2.1.3. Reposting Network

Reposting network, based on the reposting and commenting behaviors of users, has combined the features of the first two networks. It supports researchers to study community relationships from the perspective of information propagation. It can be also used to analyze the information propagation process among users. In this network, the node $v_i \in V$ represents a user and the edge $e_k = (v_i, v_j) \in E$ means the user v_j reposts or comments the message posted by the user v_i.

The dynamic feature is one of the important features of the reposting network. The basic task is to identify how information diffuses across multiple groups of users. Cui et al. [CWL*14] present a novel approach, GraphFlow, which examines and analyzes dynamic graphs based on the summarization of the graph metrics. It offers a static flow visualization for the reposting network to show the structural changes over time. Wang [WLC*16] et al. generalize the messages, topics as ideas and propose IdeaFlow, which analyzes the propagation of a set of correlated ideas among several pre-defined groups. Their work addresses the lead-lag patterns in the diffusion process. These works present a good overview for the dynamic diffusion process. To analyze the diffusion process of a central user in details, D-Map [CCW*16], proposed by Chen et al., explores information diffusion among social communities based on the reposting networks in Sina Weibo. Their reposting networks are constructed by merging the same users in the original diffusion network of messages. D-Map shows the propagation of information based on a map metaphor (Figure 4e). The technique provides a clear and intuitive visual summary of the dynamic ego-centric diffusion process.

Besides, general-purpose entities can be derived from NLP (Natural Language Process) techniques from reposting networks. The entities include brands, person names, products, etc. Yuan et al. [YWL*14] propose a system to identify key players' names and explore their roles in the reposting network. The system uses a multi-faceted filter to enable exploring statistics of the additional entities, including location, time, followers, etc., derived from NLP techniques. To further analyze the entity's behavior, Cao et al. [CSL*16] propose TargetVue, which detects anomalous users via the Time-adaptive Local Outlier Factor (TLOF, a machine learning model) and visualizes the suspicious users by summarizing user's communication activities, features and social interactions with others (Figure 4f).

In short, social network analysis intrinsically adopt the visualization techniques like node-link graph [HB05, RZW*14], matrix [HFM07], and arc diagram [HF07]. For the diffusion network and reposting network, special design such as map-like [CCW*16, CSL*16] and river-like [ZCW*14, CWL*14, WLC*16] techniques are used to address the characteristics of the diffusion properties and a large number of messages and people.

2.2. Spatial Temporal Visualization

There are two main sources of spatial information on social media. First, people might indicate where they come from, e.g. their hometown or living places. Second, people can post messages with geo-tagged information, e.g. geo-tagged tweets. Using Weibos in China as an example, the ratio of geo-tagged messages is around 3% [CYW*16]. Considering a large number of social media messages generated every day, the amount of such geo-tagged messages is large.

Based on the two categories, we can derive three main research focuses for the spatial temporal analysis in the social media. First, we have people's location information. Messages are reposted by people, from which we can infer how information diffuses across different regions as well as participants' distribution. Second, social media messages may have geo-tags. Researchers can analyze the spatial temporal social event distributions based on the geo-tagged information. Lastly, the geo-tagged messages can be constructed as trajectories to infer the movement of social media users. In short, the research includes the geographic information diffusion analysis, spatial temporal event analysis and movement analysis (Figure 2).

2.2.1. Geographical Information Diffusion Analysis

Considering the geo-location information in the user profile, we can estimate that most users usually live in the city/region they mark as living places. $V = \{v_i, v_2, \ldots, v_n\}$ is a set of users. For each v_i, he/she has an attribute of h_i, which represents his/her living place. Usually, such information would not have a precise latitude and longitude value, but it indicates a region, a city or even a country. Researchers using such information can provide a country-level [CLS*12] and city-level [CCRS13, ZLW13] information diffusion analysis. Such a diffusion process integrates both spatial and temporal information, which requires in visualizing the dynamic scenarios.

Cao et al. present Whisper [CLS*12], one of the earliest visual analytics work to represent and analyze the spatial temporal information diffusion process over the world (Figure 5a). Whisper includes a sunflower visual metaphor, describing how tweets in one topic spread from the source center region to the users all over the world. Users can perceive the temporal trends, i.e., topic evolution in the spatial context of such a diffusion process. Croitoru et al. [CCRS13] present the Geosocial Gauge prototype system, which also highlights the retweeting network with the map context view. Additionally, Zhang et al. [ZLW13] provide a sentiment analysis component in analyzing such geographical reposting behaviors, with case studies on the Sina Weibo Service in China. In their work, users can understand how a topic is reposted by people from different regions and seek for the sentiment distribution in different regions for each topic. In the research work focusing on information diffusion, WeiboEvents provide the attributes view to highlight where the retweeting users are from [RZW*14]. We can observe the geographical distribution of a special topic and find different local events with the significant spatial distributions of participants in social media.

In short, the geo-location information provides the spatial context of users in analyzing the information diffusion process. However, since it is not derived from the social media message itself, it can only provide an overview as well as geographic context for such an analysis. More researchers are focusing on the geo-tagged social media analysis as introduced in the following sessions.

2.2.2. Spatial Temporal Event Analysis

The geo-tagged social media naturally provides the spatial, temporal, textual and multimedia information in the messages. With the development of smartphones and GPS technologies, it is easy to post geo-tagged messages. A spatial temporal social event is defined as a series of geo-tagged messages that correlate by similar topics in a spatial temporal context. We define the general form of a social event $S = (V, G)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of users, and $G = \{g_1, g_2, \ldots, g_m\}$ is a set of geo-tagged social media messages, $g_j = (t_j, pos_j, text_j, addAttr_j)$, and each message includes a time stamp, position (with latitude and longitude), text information, etc. Each user in V can post one or multiple messages.

In 2011, MacEachren et al. developed a system that integrates such information to observe the spatial temporal distributions of social events [MJR*11, MRJ*11]. They provide basic components such as a spatial view, a temporal view and a content view. These views are coordinated together to enable the visual exploration. For geo-tagged messages, the visualization forms vary from a point-based representation [CCRS13] to a heat map visualization [MJR*11] or density map visualization [CTJ*14] to further reduce the clutter and show aggregated information (Figure 5c). Besides 2D representation, histograms on the 3D globe are also used [KWD*13]. From the temporal perspectives of the social events, Chae et al. [CTB*12] combine a seasonal trend decomposi-

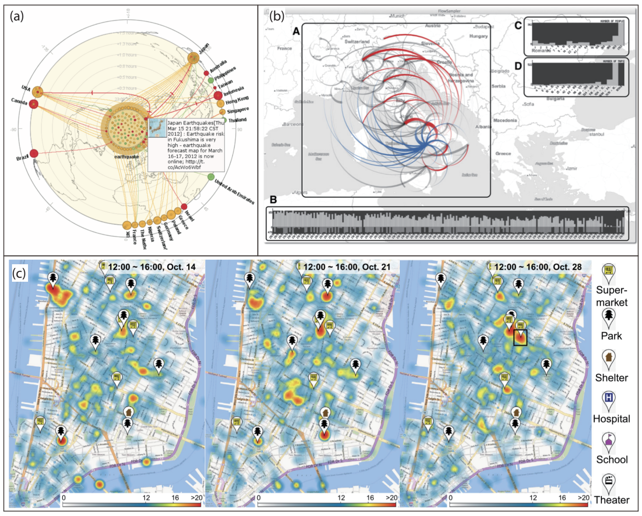

Figure 5: *Spatial temporal visualization in social media. (a) Spatial temporal circular design in Whisper, integrating the spatial temporal information and representing the diffusion process across multiple regions [CLS*12]. (b) A flow-based visualization, showing the movement among different regions [CMSVM14]. (c) A density map design with the aggregation to show the distribution of the social events [CTJ*14].*

tion function with spatial temporal visualization to investigate the normal and abnormal events. These are the basic visualization and interactions of spatial temporal visualizations for geo-tagged information.

To deeply investigate the features of spatial temporal events, advanced pre-processing, machine learning and interaction methods are fused together within the visualization techniques [TBK*12, BTH*13, TKE*15, KWD*13]. To deal with massive geo-tagged Twitter data, Thom et al. propose ScatterBlog [TBK*12], a visual analytics system which can investigate a large number of events and visualize the text distribution representing these events. In the improved version of ScatterBlog2 [BTH*13], they provide a comprehensive filtering mechanism. There are two stages of the analysis process, one is classifiers and filters building stage and the other one is the real-time monitoring based on the defined filters. Such customized interactive and learning tools could further help users identify the events of interest on the map dynamically. They share the case results for disaster management with their tools [TKE*15]. Krueger et al. also design the spatial filtering interactions for the similar goals on the map [KTE15].

Besides the visualization of events derived from geo-tagged messages, researchers fuse the geo-tagged messages with other sources of data. Prieto et al. [PHE*15] propose a circular visual design to use both land usage data and geo-tagged messages to support the analysis in urban planning. To effectively understand the urban behavior, McKenzie et al. [MJG*14] visualize the geo-tagged Twitter distribution for each position of interest (POI), and visually compare the different patterns of these distributions in the city. Thus they can understand the different activity types in different regions of the city.

2.2.3. Movement Analysis

Besides the spatial temporal aggregation of geo-tagged messages to reflect social events, we observe other aspects of insight from the trajectory perspective. For each user v_i, he/she may post a series of geo-tagged social media messages G, where $G = \{g_1, g_2, ..., g_k\}$. Chronologically, we can construct a rough trajectory from these data. Because of the large amount and wide distribution of the data, analysts and experts in social science can extend their original analysis process, which is based surveys with a limited number of people to a larger scale. However, there are reliability and uncertainty problems when dealing with such trajectory data [CYW*16].

The basic visualization of movements derived from geo-tagged messages leads to a flow representation [CMSVM14] (Figure 5b). The spatial scale of such visualization ranges from the global scale [KSB*16], via the inter-city scale [CYW*16,LSKG14] to the inner city scale [WZSL14,CCJ*15]. Researchers focus on the human mobility patterns based on these social media data. Krueger et al. visualize the global movement based on geo-tagged Twitter with density maps [KSB*16]. Liu et al. investigate the spatial interaction and distance decay of the mobility patterns [LSKG14]. They visualize and analyze the check-in social media data from a million individuals within 370 cities. Through the flow visualization with detected communities, they find that movement communities are spatially consistent with province boundaries in China. Besides the spatial pattern, Wu et al. also visualize temporal movement patterns at the inner city scale [WZSL14]. They represent the movements among POIs in the cities with a transition matrix. Also at the inner city scale, Chae et al. detect and filter common sequences of the movements and visualize the movement patterns in specific events, such as the Boston Marathon event [CCJ*15].

Andrienko et al. [AAS*12] summarize derived trajectories from social media as episodic movement data. The definition is "data about spatial positions of moving objects where the time intervals between the measurements may be quite large and therefore the intermediate positions cannot be reliably reconstructed by means of interpolation, map matching or other methods". They argue that with spatial and temporal aggregation, the general patterns can be derived [AAS*12]. They also visualize trajectories with semantics in a space-time cube [AAF*13]. Chen et al. [CYW*16] summarize the characteristics of such sparse trajectories from geo-tagged Weibos and provide a visual analytic system combining spatial, temporal and attributes aggregation to analyze movement patterns. They also propose a Gaussian Mixed Model-based uncertainty model, to guide the filtering and analyzing the process. To further explore such patterns, Krueger et al. [KSB*16] provide a comparison view for the social media trajectories with heterogeneous movement data. These works indicate that many interesting patterns can be found, by addressing the challenges of large data amounts and uncertainties.

In short, for the geographic information, map-based visualization [CLS*12, CCRS13, ZLW13], density-based visualization [TBK*12, BTH*13, TKE*15, CTJ*14] and timelines [KSB*16] are heavily used, to address the spatial temporal characteristics of the data. Besides, wordles are used to represent the text distribution on the map [BTH*13]. For the trajectory, aggregated and individual flows are visualized on the map [CYW*16, CMSVM14].

2.3. Text Visualization

One critical component of social media is the content. Users generate contents, including text, images, multimedia, etc., to share information, give opinions, spread news and connect to other people, etc. Doerk et al. provided a visualization system for social media content - Visual BackChannel [DGWC10] in 2010. The system is among the earliest to visualize the main parts of social media content. They use a timeline to show the keywords and topics, a circular view to indicate the participating people, and a text list to show the raw data and an image clouds. They also provide a search form and interactive filters, allowing users to explore the discussing themes and content based on users' reposting behaviors.

We summarize three types of focus, including keywords, topics and sentiments visualization and analysis in an incremental manner. According to the definitions in [CC16, LYW*16], words and topics are two different levels to reveal content. In the context of social media messages, keywords are the words with high mentioning frequency. Visualizing keywords in the social media extracts word-level semantics. Topics are the summarized subjects from social media content. Visualizing topics in the social media extracts topic-level semantics, which are highly summarized and derive the themes of contents. Sentiments are summarized from contents with the attitude of social media users.

2.3.1. Keywords Visualization

Words are the foundations of text. One characteristic of social media text is that the text length in the messages is limited and usually short. Keywords extracted from the text can basically represent the overall meaning of the content. WordCloud [VWF09] is a common approach. A group of words is laid out on the plane, with the size indicating the frequency and importance. However, simply using word clouds can not reveal deep insights from social media contents. Archambault et al. [AGCH11] propose ThemeCrowds, using a hierarchical visualization to investigate keywords on different levels and layout these keywords chronologically (Figure 6a). It enables the detailed investigation of keywords and allows users to explore, search and filter keywords of interest.

To have a deeper understanding of social media keywords, further research addresses dynamic features of keywords [BBD*12] and the relationship among keywords [HWS17, LLZ*16]. Best et al. [BBD*12] propose a web-based visualization system, visualizing the dynamic keywords streams derived from streaming data from Twitter. The visualization form adopts the ThemeRiver metaphor [HHN00] and dynamically update the text along time. It allows convenient monitoring and the detection of social events. Moreover, the relationship among keywords can be further investigated to understand social media. Hu et al. [HWS17] propose SentenTree, visualizing the frequent keywords co-occurrence patterns in the sentences of social media messages. It can help users quickly understand the concepts and opinions of a large social media text collection. Besides the keywords correlation, Liu et al. [LLZ*16] summarize the correlation of keywords, users, and hashtags in Twitter. They propose an uncertainty model to retrieve semantic information, important keywords and users. Wanner et al. [WJS*16] identify interesting financial time series intervals and corresponding news feature by extracting keywords in the news and social media. Their system supports users to analyze the relationships between financial patterns and text. Beyond these works addressing on the keywords level, researchers are also interested in the topics that derived from the text information in social media.

2.3.2. Topics Visualization

On social media platforms, people discuss and share their opinions about specific events. These events could be posted by a news agency, a famous person, or a witness of events in the real world. Multiple topics are usually generated from social events [DWS*12], and change along the time across multiple groups of people [CLWW14, WLY*14]. We observe two main themes in the research of topic visualization. One is addressing the hierarchical feature of topics, and the other is identifying the interactions among topics.

Dou et al. [DWS*12] define social events, including four elements as the topic, time, user and location. Based on these, they derive multiple topics from the event and visualize them with parallel rivers (Figure 6b). The fluctuation of the river indicates the number of Tweets within a specific topic. Based on these features, they can analyze the reasons why events break out and identify the sources and related topics [RWD14]. Wang et al. extend the system into a new one, I-SI, which improve the scalability when analyzing large groups of topics in the events [WDM*12]. Though these systems work well for detecting and visualizing the topic distribution along the time, the flatness of the topic definition limits their coverage. To solve this problem, Dou et al. raise a HierarchicalTopics visualization system [DYW*13], which uses the Topic Rose Tree to calculate the topic hierarchies. They also extend the visualization form with a tree representation with the detailed topic shown in the topic river. The proposed method indeed enables a detailed exploration in the multi-level form and expands the exploration space. However, it still has the limitation that these hierarchies are calculated for the whole time ranges in the preprocessing stage. Cui et al. point that the topic hierarchies are also dynamically changes in the real scenarios [CLWW14]. Thus, they propose the RoseRiver, which uses the Tree-Cut algorithm to detect different topic hierarchies along the time. Their method can adaptively find the suitable hierarchical levels for the different time and visualize them in the river metaphor. These are research works focusing on the hierarchical features of the topic visualization.

The interaction and influences among multiple topics are of interest to researchers. Because of the transition of social media users' focus, the topic they evolve will dynamically change. However, the traditional methods can not represent such changes [XWW*13, WLY*14]. Xu et al. first visualize the competition behaviors among topics in the social media with a river metaphor [XWW*13]. Users can easily see how topics emerge, replace others and die along the time (Figure 6c). One step further, Sun et al. find that relationships among topics are not only com-

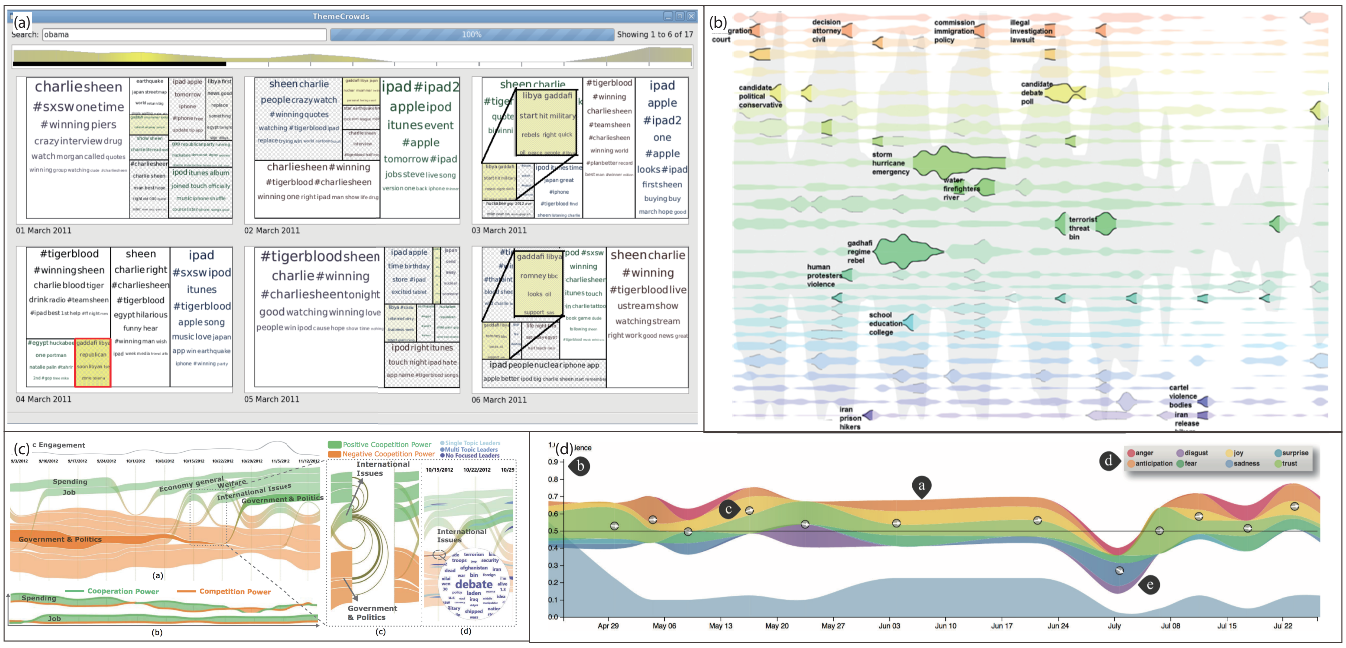

Figure 6: *Keywords, topic and sentiment visualization with a river-based metaphor. (a) ThemeCrowds, a multi-level dynamic word clouds visualization [AGCH11]. (b) Leadline, visualizing the dynamic event and topic evolution [DWS*12]. (c) Topic competition visualization, representing how different communities' behaviors lead to the increase and decrease of the topics [XWW*13]. (d) Sentiment visualization, showing the eight extracted sentiment along the time [ZGWZ14].*

petition but also collaboration [SWL*14]. Based on this, they propose EvoRiver, in order to visualize the collaboration-competition ("Coopetition") relationship of topics. They use color to encode the tendency of collaboration or competition. By extending the stack graph, they visualize the distribution of the topics and key players on the coopetition river. Considering the people behavior together with the evolution of the topics, Wu et al. propose OpinionFlow [WLY*14], observing how people diffuse the information in different topics. We can observe that the intrinsic dynamic feature leads many researchers to use the river metaphor to visualize topic distribution and evolution.

2.3.3. Sentiments Visualization

Sentiment is behind most messages people post. By summarizing the public sentiment towards social events, researchers can estimate the general public's attitude for better understanding social events. These types of research are essential in many application domains, such as politics, advertisement, etc.

To visualize sentiment, Hu et al. [HYZ*13] use a matrix visualization to encode supporting and opposite opinions with yellow and green color. Further, Zhao et al. [ZGWZ14] visualize a deeper classification of sentiment (Figure 6d). It follows the Discrete Categorical Model [Plu01], which includes four pairs of sentiment: anger - fear, anticipation - surprise, joy - sadness, and trust - disgust. The personal emotion information is visualized along the stack graph. Compared with the personal emotion analysis, Steed et al. [SDB*15] visualize the aggregated emotional dynamics of a large group of people with the high-dimensional projection.

Rohrdantz et al. [RHD*12] propose automatic methods and interactive visualizations to extract sentiment from text document streams. The system supports users to analyze sentiment patterns, explore time-stamped customer feedback and detect critical issues. To summarize the users' requirements in sentiment analysis, Kucher et al. [KSBK*15] find that users are not only interested in the sentiment or mood, but also the attitude towards the events. They summarize the attitude and sentiment changes with the timeline, together with the text analysis system to support the tasks. Besides the general methods, Liu et al. present SocialBrands for brand managers to analyze public perceptions of brands in social media. It focuses on the special name entities as brands [LXG*16]. They propose a circular design supporting visual comparison of multiple perception matrix.

In short, for text data, wordles [SWL*14, AGCH11] and river-like visualizations [XWW*13, WLY*14, SWL*14, ZGWZ14] are usually used to understand the text and dynamic behaviors in social media.

2.4. Visualization Techniques Summary

We extract 17 visualization techniques from the selected 68 papers (Figure 7). The methodology is that we extract one or multiple main techniques in each paper and calculate the statistics. We try to detect general techniques which can inspire future work in two aspects. First, readers know which techniques can be used for specified data types. Second, users can further extend and combine multiple techniques depending on the targeted problems. In the fol-

Category	Node-link Diagram	Matrix	Arc Diagram	Small Multiples	Glyph	3D Representation	Timeline	Tree	Space-Filling	Parallel Coordinates	Map-based Visualization	Flow on the Map	Map-like Visualization	River-like Visualization	Density-based Visualization	Circular Design	Wordle
N1	2 [HFM07]; [CZH*09]	2 [HF07]; [HFM07]	2 [HF07]; [BCD*10]	2 [BCD*10]; [BN11b]	1 [BN11b]	2 [CZH*09]; [BCD*10]	1 [CZH*09]	0	0	0	0	0	0	0	0	0	0
N2	1 [RZW*14]	0	0	0	0	0	1 [RZW*14]	1 [VWH*13]	2 [VWH*13]; [ZCW*14]	1 [LQC*13]	0	0	1 [LWYZ15]	1 [ZCW*14]	0	0	0
N3	1 [CSL* 16]	0	0	0	1 [CSL* 16]	0	2 [CWL*14]; [WLC* 16]	0	0	1 [YWL* 14]	0	0	2 [CCW* 16]; [CSL* 16]	2 [CWL*14]; [WLC* 16]	1 [CWL*14]	0	0
G1	1 [CLS* 12]	0	0	0	2 [CLS* 12]; [ZLW13]	0	0	0	0	0	3 [CLS* 12]; [CCRS13]; [ZLW13]	2 [CLS* 12]; [ZLW13]	0	0	0	1 [CLS* 12]	1 [CCRS13]
G2	0	0	0	1 [MJG* 14]	2 [BTH* 13]; [TKE* 15]	1 [KWD* 13]	9 [MJR* 11]; [MRJ* 11]; [CTB* 12]; [KWD* 13];[CTJ* 14]; [TBK* 12]; [BTH* 13]; [TKE* 15]; [MJG* 14]	0	0	0	11 [MJR* 11]; [MRJ* 11]; [CTB* 12]; [KWD* 13]; [CTJ* 14]; [TBK* 12]; [BTH* 13]; [TKE* 15]; [PHE* 15]; [MJG* 14]; [XSX* 14]	0	0	0	8 [MJR* 11]; [MRJ* 11]; [CTB* 12]; [CTJ* 14]; [TBK* 12]; [BTH* 13]; [TKE* 15]; [MJG* 14]	1 [PHE* 15]	5 [CTB* 12]; [KWD* 13]; [TBK* 12]; [BTH* 13]; [TKE* 15]
G3	0	0	0	3 [CYW*16]; [WZSL14]; [AAF* 13]	0	1 [AAF* 13]	1 [CYW*16]	0	0	1 [CYW*16]	8 [CYW*16]; [CMSVM14]; [CCJ* 15]; [KSB* 16]; [LSKG14]; [WZSL14]; [AAS* 12]; [AAF* 13]	7 [CYW*16]; [CMSVM14]; [CCJ* 15]; [KSB* 16]; [LSKG14]; [AAS* 12]; [AAF* 13]	0	0	4 [CYW*16]; [CCJ* 15]; [KSB* 16]; [LSKG14]	1 [CYW*16]	1 [CYW*16]
T1	1 [LLZ* 16]	0	0	2 [AGCH11]; [WJS*16]	2 [LLZ* 16]; [WJS*16]	0	6 [DGWC10]; [AGCH11]; [BBD* 12]; [DNYKS11]; [DNK10]; [WJS*16]	1 [HWS17]	0	0	0	0	0	2 [DGWC10]; [BBD* 12]	0	1 [DGWC10]	5 [AGCH11]; [HWS17]; [DNYKS11]; [DNK10]; [WJS*16]
T2	1 [WLY*14]	0	0	4 [DWS* 12]; [RWD14]; [DYW*13]; [DCE* 15]	1 [DCE* 15]	0	8 [DWS* 12]; [RWD14]; [WDM*12]; [DYW*13]; [CLWW14]; [XWW*13]; [WLY*14]; [SWL*14]	1 [DYW*13]	1 [LWW* 13]	1 [DCE* 15]	0	0	2 [GHN13]; [LWW* 13]	7 [DWS* 12]; [RWD14]; [WDM*12]; [CLWW14]; [XWW*13]; [WLY*14]; [SWL*14]	1 [WLY*14]	0	5 [DWS* 12]; [CLWW14]; [XWW*13]; [SWL*14]; [DCE* 15]
T3	1 [CLL* 14]	2 [HYZ*13]; [RHD*12]	0	3 [LWM14]; [LKT* 14]; [LXG*16]	2 [CLL* 14]; [LXG*16]	0	10 [ZGWZ14]; [SDB*15]; [KSBK*15]; [LWM14]; [LKT* 14]; [BMZ11]; [MBB* 11]; [CLL* 14]; [RHD*12]; [LXG*16]	0	0	1 [LKT* 14]	1 [MBB* 11]	0	0	1 [ZGWZ14]	1 [SDB*15]	1 [LXG*16]	3 [ZGWZ14]; [LWM14]; [LKT* 14]

Figure 7: *Summarized 17 general visualization techniques used in social media visualization for nine data categories. One or multiple techniques can be derived from one paper.*

lowing sections, we also discuss how we can combine visualization and analytic methods to solve problems with social media data.

3. Visual Analytics Techniques

Considering the complex characteristics of social media data, more visual analytics methods combining the above specific visualization techniques and mining algorithms are proposed. In this section, we first analyze the compounded visualization techniques and categorize the relationships of multiple techniques used in these works. Following this, we summarize the research goals of selected visual analytics papers in social media and illustrate how these techniques work for specific analytical goals.

3.1. Methodology

As mentioned in the above sections, we have marked each paper with one main category and multiple other categories. To further summarize visual analytics combining multiple techniques, we summarize the concurrent entities with corresponding visualizations used in the same work and build up a matrix (Figure 8). Rows and columns represent the categories, and the number in the cell indicates how many papers combine the two visualization techniques.

There are two ways to summarize such concurrent relationships for calculating the matrix. One is to identify the concurrent relationship based on the main category and other categories. The other approach is to build up a complete graph based on all the categories the paper share. It can be regarded as a star shape relation and a complete graph relation. For example, when considering Wang et al.'s work [WLC*16], we mainly categorize it as investigating "Reposting Networks" with a river-based visual metaphor. At the same time, they use geographic information for diffusion analysis, keywords and topic analysis to support the proposed method. In such situation, adding the concurrent relationships of geographic information diffusion with keywords or topics seems not appropriate because they are not mainly addressed. In our review, we emphasize the pair-wise relationships. We choose to add the concurrent relationships between the main category (diffusion network analysis) and other categories (geographic information diffusion, keywords, and topics). We define the categories in the row as the main categories, and the categories in each column as other categories. The matrix is not symmetric and the values in the diagonal are the number of papers with the specified main category.

Besides the pair-wise relationship, we are also interested in the many-to-many relationship among the visualization techniques. Thus, we use a graph-based category visualization method to observe the patterns (Figure 9). In the visualization, there are two types of nodes. The circle nodes represent categories and the rectangle nodes represent the intersections of multiple categories. The links connecting rectangles and circles indicate the papers sharing the connected categories. The size of the nodes encodes the paper count. With the visualization, we can see the paper number of each category as well as the concurrent relationship among papers and categories. With these results, we can lay the foundations for analyzing the different goals of visual analytics works of social media.

After calculating the correlations among nine categories of entities and corresponding visualization techniques, we summarize the general goals for social media visual analytics (Figure 3). One paper can achieve one or multiple goals. We further collect the visual analytics pipelines of each paper and summarize the general visual analytics pipeline for social media (Figure 11), with six important goals. By aggregating the main categories of the investigating entities to the six visual analytics goals, we build up the visual analytics matrix (Figure 10). With the matrix and the correlation analysis, we can propose hypotheses about the general research trends of social media visual analytics.

3.2. Relationship among Visualization Techniques

Based on the category distribution of each paper (Figure 3), we collect pair-wise relationships of the entities and their visualizations (Figure 8) and the complete relationships among categories and the entities (Figure 9). Generally, we can see keywords and topics analysis are widely used in many visual analytics papers addressing social media. The amounts are 48 and 30 respectively (Figure 3).

Category	Follower Network	Diffusion Network	Reposting Network	Geograhic Information Diffusion	Spatial Temporal Event Distribution	Movenment Trajectory	Keywords	Topic	Sentiment
Follower Network	5 [HF07];[HFM07]; [CZH*09]; [BCD*10];[BN11b]	0	0	0	0	0	1 [CZH*09]	0	0
Diffusion Network	0	5 [VWH*13]; [RZW*14]; [LQC*13]; [LWYZ15]; [ZCW*14]	0	1 [RZW*14]	0	0	1 [RZW*14]	2 [LWYZ15]; [ZCW*14]	2 [RZW*14]; [ZCW*14]
Reposting Network	1 [YWL*14]	2 [CCW*16]; [YWL*14]	5 [CWL*14]; [CCW*16]; [WLC*16]; [CSL*16]; [YWL*14]	1 [WLC*16]	0	0	4 [CCW*16]; [WLC*16]; [CSL*16]; [YWL*14]	2 [WLC*16]; [CSL*16]	1 [CSL*16]
Geograhic Information Diffusion	1 [CCRS13]	2 [CLS*12]; [ZLW13]	2 [CLS*12]; [CCRS13]	3 [CLS*12]; [CCRS13]; [ZLW13]	2 [CLS*12]; [CCRS13]	0	1 [CCRS13]	0	2 [CLS*12]; [ZLW13]
Spatial Temporal Event Distribution	0	0	0	0	11 [MJR*11]; [MRJ*11]; [CTB*12]; [KWD*13]; [CTJ*14]; [TBK*12]; [BTH*13]; [TKE*15]; [PHE*15]; [MJG*14]; [XSX*14]	0	9 [MJR*11]; [MRJ*11]; [CTB*12]; [KWD*13]; [TBK*12]; [BTH*13]; [TKE*15]; [MJG*14]; [XSX*14]	7 [CTB*12]; [KWD*13]; [CTJ*14]; [TBK*12]; [BTH*13]; [TKE*15]; [MJG*14]	0
Movenment Trajectory	0	0	0	0	6 [CYW*16]; [LSKG14]; [WZSL14]; [CCJ*15]; [AAS*12]; [AAF*13]	8 [CYW*16]; [CMSVM14]; [KSB*16]; [LSKG14]; [WZSL14]; [CCJ*15]; [AAS*12]; [AAF*13]	2 [CYW*16]; [CCJ*15]	1 [AAF*13]	0
Keywords	0	2 [BBD*12]; [LLZ*16]	2 [DGWC10]; [LLZ*16]	0	0	0	9 [DGWC10]; [DNYKS11]; [DNK10]; [FS14]; [AGCH11]; [BBD*12]; [HWS17]; [LLZ*16]; [WJS*16]	4 [DGWC10]; [DNYKS11]; [DNK10]; [BBD*12]	3 [DNYKS11]; [DNK10]; [FS14]
Topic	0	3 [XWW*13]; [WLY*14]; [SWL*14]	1 [WLY*14]	1 [WDM*12]	2 [DWS*12]; [WDM*12]	0	11 [DWS*12]; [DCE*15]; [GHN13]; [LWW*13]; [RWD14]; [WDM*12]; [DYW*13]; [CLWW14]; [XWW*13]; [WLY*14]; [SWL*14]	11 [DWS*12]; [DCE*15]; [GHN13]; [LWW*13]; [RWD14]; [WDM*12]; [DYW*13]; [CLWW14]; [XWW*13]; [WLY*14]; [SWL*14]	1 [WLY*14]
Sentiment	0	0	1 [CLL*14]	1 [BMZ11]	2 [SDB*15]; [BMZ11]	0	11 [HYZ*12]; [ZGWZ14]; [SDB*15]; [KSBK*15]; [LWM14]; [LKT*14]; [BMZ11]; [MBB*11]; [CLL*14]; [RHD*12]; [LXG*16]	2 [HYZ*12]; [LXG*16]	11 [HYZ*12]; [ZGWZ14]; [SDB*15]; [KSBK*15]; [LWM14]; [LKT*14]; [BMZ11]; [MBB*11]; [CLL*14]; [RHD*12]; [LXG*16]

Figure 8: *Pair-wise concurrent relationships of the entities in social media visual analytics. The number indicates the concurrent times the two entities are used in the same paper. It also reflects the relationships of the corresponding visualization techniques.*

They are widely used in analyzing network and geographic information, which reflects the feature of user-generated content in the social media. Different from the traditional social network and spatial temporal analysis, integrating keywords, topics and even sentiment analysis into the visual analytics process can help users gain understanding of the semantics of the data.

3.2.1. Inner Category Patterns

With the overall understanding of the techniques trends, we drill down to the details of the pair-wise patterns. First, we check the inner category patterns of each of the three main categories.

In network visualization, we first observe that the visual analytics works investigating follower networks in social media seem to be isolated. As discussed in Section 2.1, these works mainly propose new visualization techniques addressing improving the visual representation [HF07,HFM07], dimension analysis [BCD*10], etc. Especially in recent years, there are few works addressing the follower network analysis. Besides, the diffusion network and reposting network are correlated. The two networks are the result of people's posting and reposting behaviors. The difference is that the diffusion network focuses on the message spreading while the reposting network addresses the people's relationship based on their interactions. Currently, there are two works explicitly addressing both. As an example, D-Map derives the reposting network from the message diffusion network in Sina Weibo [CCW*16] (Figure 4e). We see the potential to tightly couple these two kinds of networks and analyze the detailed information patterns.

For the geographic information, we can summarize that the spatial temporal event distribution plays a bridging role. Works on geographic information diffusion focus on the geo-location of the posting people, while the works on spatial temporal event distribution and trajectories are based on the geo-tagged messages. The connections are that people might possibly stay in where he/she is from and post social media messages there. Whisper considers the reposting network with geographic information [CLS*12] (Figure 5a), which is a representative among a few works in both geographic information diffusion and spatial temporal event distribution. While many other works either visualize the geo-tagged social media messages distribution [MJR*11, TBK*12] or connect the geo-tagged social media messages of the same people to construct trajectories [CWLY16, KSB*16]. Event distribution and trajectories are usually combined for analysis. By aggregating the geo-tagged messages on the map, users can understand the event distribution. Further by connecting the same user's sequence of messages, users can identify the dynamic movement patterns and correlate the analysis of the event distributions [CYW*16, LSKG14, CCJ*15].

The text related research – keywords, topics and sentiment – are largely correlated. It means that they are usually combined in the analysis. We can see the different levels to derive the semantics, from the word level to topic level, and finally to the sentiment level. Works which use topic analysis usually take keyword level analysis first [DYW*13] or visualize the keyword distribution among each topic [XWW*13, WLY*14, SWL*14]. In most scenarios, the keywords level analysis also provides foundations for sentiment analysis [ZGWZ14].

3.2.2. Inter Category Patterns

Visual analytics makes use of multiple entities and visualization techniques. Besides the widely used semantic related information mining and visualization, we also observe many interesting patterns from the matrix (Figure 8) and the graph (Figure 9). In this discussion, we provide a case-by-case study to shed light on interesting inter-category patterns.

First, we can see there are five papers in total that concurrently conduct geographic information diffusion analysis and network analysis (Figure 8). It is easy to understand because it is a special case of the diffusion network and reposting network. There are two types of such correlation. WeiboEvents shows the participating people's geographic information distribution [RZW*14]. In the other case, ideaFlow [WLC*16] directly provides a special case of information diffusion across multiple continents.

Second, we find that seven papers simultaneously analyze the spatial temporal event with both keywords and topic analysis. The triple connection is shown with the largest rectangle in the graph (Figure 9). We see that it has become common to project

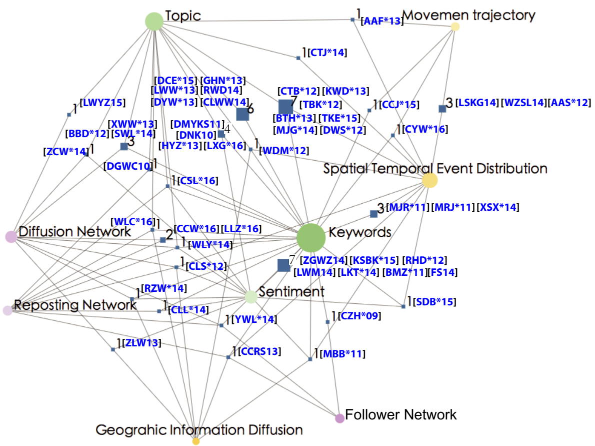

Figure 9: *A node-link visualization to show the concurrent correlations among different categories. A circle represents the category while a rectangle represents the intersection of categories. The size encodes the paper count inside. The link shows the intersection connections.*

Category	N1	N2	N3	G1	G2	G3	T1	T2	T3
Visual Monitor - 30	0	7	4	3	12	1	24	16	9
Feature Extraction - 43	6	10	8	2	9	8	27	16	12
Event Detection - 22	2	5	5	5	11	0	19	10	8
Anomaly Detection - 12	0	2	1	1	5	2	8	7	5
Predictive Analysis - 7	0	1	0	1	4	2	4	2	2
Situation Awareness - 6	0	2	3	0	3	0	6	4	2

Figure 10: *The matrix of visual analytics goals and the targeting entities. There are six summarized categories of different visual analytics goals and nine categories of entities. From the matrix, we can make hypotheses of the suitable entities and visualization methods for different analytical goals.*

the event density on the map with the derived semantic information [BTH*13, TKE*15]. However, compared with it, the movement analysis used much less semantic techniques, only with some keyword-level analysis. From this point and collected research, we can make hypotheses on the research challenges in analyzing both spatial temporal movement patterns and semantics in the same time. If such analysis can be achieved, the movement with clear semantics will have impacts on many topics of research in social science [CYW*16].

Third, we see the research trends on analyzing the different diffusion patterns with semantics in the diffusion network. From the topic distribution [BBD*12], topic evolution [DYW*13, SWL*14], to specific patterns of topic competition and collaboration [XWW*13, SWL*14], researchers explore the diffusion network from deeper semantic evolutions. However, we can see there are not too many works addressing the sentimental changes in the diffusion network or reposting network. Though challenging, understanding how people's attitude change towards an event and how such sentiment changes among people's network would be important for analyzing social media.

In short, we can find many interesting inter-categories and inner categories patterns from the collected researches. Next, we will summarize the visual analytics goals from the correlations among the categories, to illustrate research trends and challenges in this area.

3.3. Visual Analytics Categories

The general visual analytics process has been summarized by [TC05, KAF*08]. It usually integrates mining methods, visualization, and interaction. According to the selected papers in social media visual analytics, we summarize them into six categories, including visual monitoring, pattern extraction, event detection, anomaly detection, predictive analysis and situation awareness (Figure 3). The general process includes the social media data processing, pre-analysis (or layout, entity extraction, trajectory construction, etc.), visualization, interactions for human-in-the-loop analysis (Figure 11).

We summarize these goals with different levels of analytical reasoning. Generally, we have a pre-request for each category. With basic visualization and interaction, visual analytics systems can provide visual monitoring with interactive functions to support details filtering (Figure 11a). By examining the proposed visual patterns and mining results, users can extract desired features within the visualization (Figure 11b). With location, time, people and text information, users can derive and identify the events visually (Figure 11c). With the visual patterns, users can classify normal and anomaly behaviors (Figure 11d). With the understanding of existing behaviors, visual analytics system can either predict future trends, based on the trained model (Figure 11e), or combine all the mentioned techniques to derive insights for situation awareness (Figure 11f). Together with the summarized pipeline and categories with different entities (Figure 10), we further discuss the analytics goals and solutions in details.

3.3.1. Visual Monitoring

The motivation for visual monitoring is to gain a quick overview of the monitored targets and it provides the basics to further identify patterns and outliers. Real-time visual monitoring is an important area with a real-time data stream [FS14, BBD*12]. Animation techniques are quite often used [MJR*11, CLS*12, CCW*16, DGWC10]. To achieve the goal of visual monitoring, a visual analytics system usually needs to prepare an overview design and provide dynamic updates. The follower network in social media does not change frequently, so there are only few works on it. In the diffusion network and the reposting network, Whisper uses the flower metaphor as an overview and dynamically updates the diffusion process from center out [CLS*12]. In their pipeline (Figure 11a), an online layout algorithm is proposed considering the topic, user group, and diffusion pathway to provide an overview for the monitoring. Compared with the network visualization, more works in spatial temporal event distribution and topic analysis provide the capability to enable visual monitoring. ScatterBlogs provide a monitoring system showing the keywords distribution on the

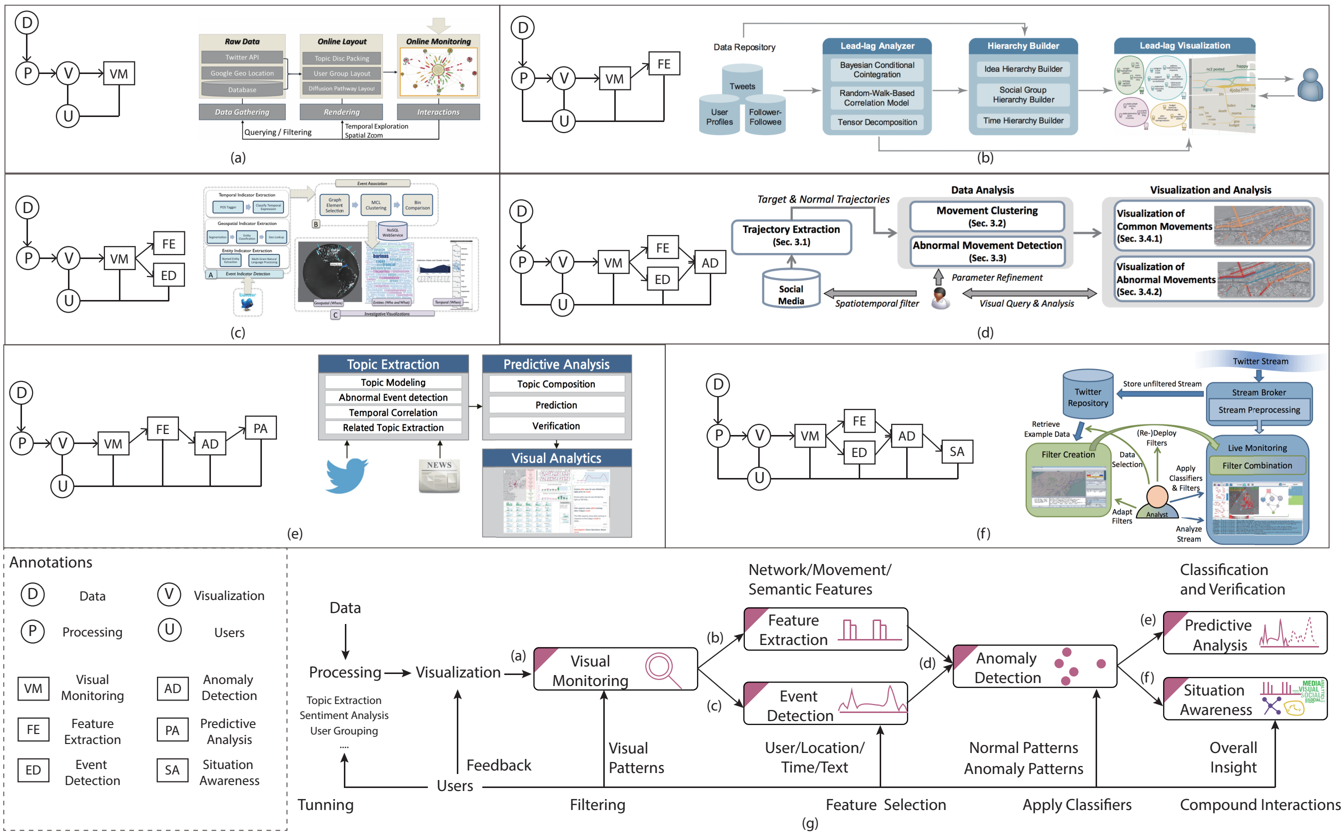

Figure 11: *Collection of six representative visual analytics pipelines. (a) Visual monitoring pipeline for geographic information diffusion [CLS*12]. (b) Extracting lead-lag patterns in ideaFlow [WLC*16]. (c) Event detection process in visual analytics [KWD*13]. (d) Anomaly detection from geo-tagged social media trajectories [CCJ*15]. (e) Predictive analysis for future event patterns through topic composition [YJ15]. (f) Visual analytics for situation awareness integrating spatial, temporal and textual information [BTH*13]. (g) Summarized visual analytics pipeline of the above six goals.*

map dynamically [TBK*12, BTH*13, TKE*15]. Besides, a river-based metaphor (Figure 6) is naturally suitable for dynamically showing the information for monitoring. For example, EvoRiver provides the visual monitoring capability for topic cooperation and competition [SWL*14].

In short, a circular visualization design [CLS*12] is a representative technique for monitoring the spatial temporal information diffusion. Besides, river-like visualization technique and timeline chart are naturally suitable for monitoring the dynamic streaming data [SWL*14]. Lastly, the wordle technique is also frequently used in visualizing and monitoring the keywords dynamics and distribution on the map [TKE*15].

3.3.2. Feature Extraction

Feature is a general term and used in many visual analytics applications. Broadly speaking, all visual analytics systems extract features. To narrow down the scope, we define a feature as the significant characteristics of one or multiple attributes in the entities of social media. It can be a range of meaningful values, special types of behaviors, etc. We mark papers into this category if they claim contributions as finding or analyzing of important features in social media. For example, ideaFlow investigates the lead-lag patterns in information diffusion [WLC*16]. In their pipeline (Figure 11b), they propose a lead-lag analyzer and a hierarchical visualization to identify and visualize the lead-lag patterns.

Besides diffusion features, there are many visual analytics works addressing spatial temporal features and semantic related features [MJG*14, CYW*16, CLWW14]. Chen et al. derive reliable movement patterns with semantics [CWLY16] (Figure 12a). They combine multiple techniques for semantic features in movement pattern analysis. Their techniques include a circular design, map-based visualization, map-based flow, attribute matrix, parallel coordinates and small multiples. Cui et al. extract dynamic hierarchical topic evolution patterns [CLWW14]. One step further, we find researchers propose visual comparisons for extracted features [CCW*16, CTJ*14, KSB*16]. For example, TravelDiff provides a visual comparison analytics for extracted movement patterns in different spatial and temporal scales [KSB*16].

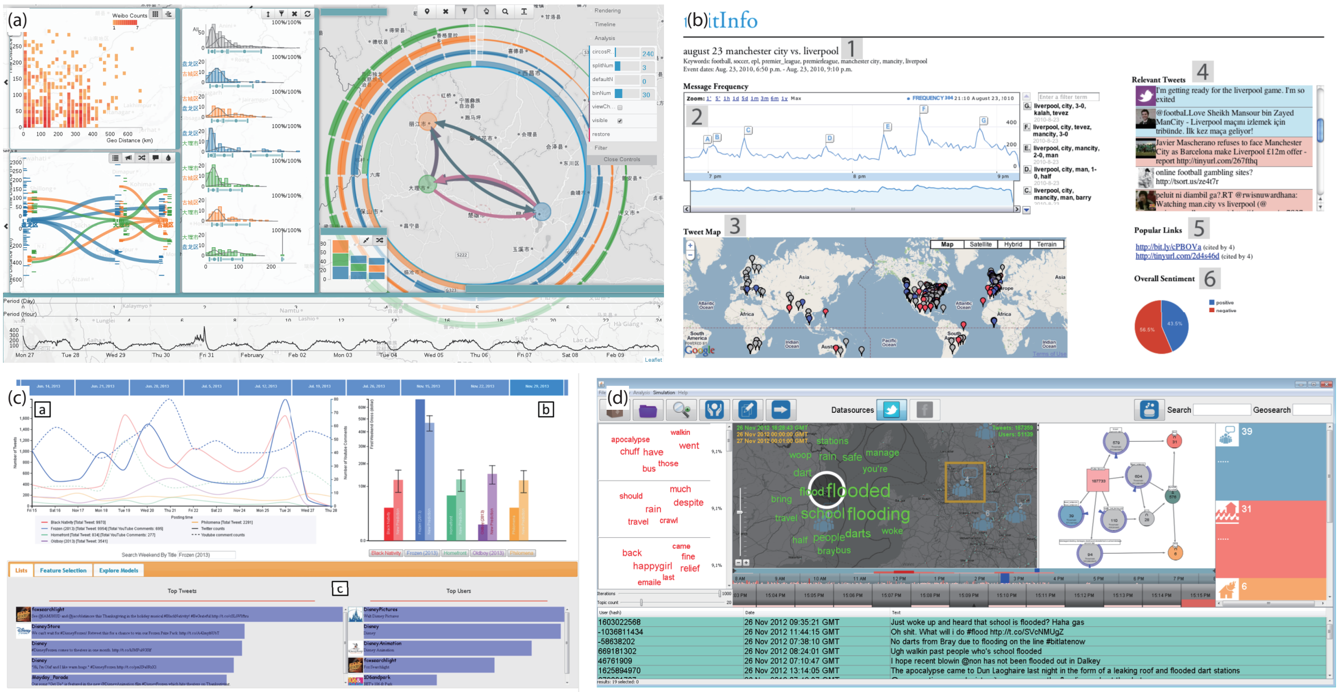

Figure 12: *Selected visual analytics systems. (a) Visual analytics system supporting the movement trajectories analysis [CYW*16], extracting the movement features in spatial distribution, time interval, temporal periods, etc. (b) Visual analytics with event detection, enabling users to explore the events in a timeline [MBB*11]. (c) Predictive analysis of box office earnings, allowing users to flexibly control the parameters [LWM14]. (d) Disaster analysis with social media to achieve situation awareness, supporting customized filter, classification design, and interaction designs [BTH*13].*

3.3.3. Event Detection

Social media can quickly reflect and affect events in the real world. Dou et al. provide a thorough discussion of event definition [DWS*12]. In [DWS*12], they define the event with four attributes $< Topic, Time, People, Location >$. It describes "When did an event start and end? What was the event about? Who was involved? And finally where did the event occur?" They also mention some previous general definitions of an event, e.g. "a noteworthy happening and a social occasion or activity" by Merriam-Webster [Eve12]. Their perspectives are from text and topic detection. Besides, in the paper collection, we also find there are many event detection works in spatial-temporal analysis. In their works, they find special events with time peaks of the social media message amount [MBB*11] in the time distribution. Moreover, in analyzing the reposting network and information diffusion network, researchers target significant events and identify how events diffuse and propagate in social media [WLC*16].

After summarizing these definitions, we go one step further for the definition of an event in our survey in the social media context. An event is defined as <*People, Messages, Time, [Reposting, Location, Themes*]>. The last three features are optional for some events. We highlight the message and the reposting relationship because events usually are exposed to the public by the posting and reposting behaviors of the people. The location information derived from geo-tagged messages might reflect a special event like the Boston Marathon [CCJ*15]. The behaviors trigger the information diffusion, and reflect the event evolution. Finally, we use Theme to represent the three levels of semantic extraction of the text. Based on the definition, we summarize two general types visualization techniques for event detection work. One is to identify the information diffusion process – the node-link diagram technique is used to show the connection and a timeline chart with a river metaphor is used to show the dynamic patterns [LQC*13, DYW*13, SDB*15]. The other one is dealing with spatial temporal events. After the events are detected with processing steps, a map-based visualization and timeline chart are used [MBB*11, BTH*13] (Figure 12b,d).

In the pipeline of Kraft et al.'s work [KWD*13], we can see that the entity, spatial and temporal information extraction are critical in the pre-process steps (Figure 11c). How to correlate and visualize this information from the general data is critical in event detection. Analyzing event evolution with details should be a challenging but meaningful research task. For additional features of location, movement and topic, Chae et al. use a seasonal-trend decomposition to detect events along the time [CTB*12]. In their following works, they detect and trace the event to enable public evaluation based on the spatial temporal distribution of the Twitter data with filtered messages [CTJ*14] (Figure 5c). Based on our extension of the event definition, we find limitations in current works. Currently, existing works have successfully detected events in either topics, location or other pre-defined contexts. However, there is lacking detailed and in-depth analysis of the detected events, which helps researchers to understand how events evolution happen and who are the key players, what are the key themes' changing patterns. This is important because we need to not only find the event, but also analyze the event in-depth.

3.3.4. Anomaly Detection

The normal pattern, in most cases, is equivalent to the common pattern, meaning that the data lies around the average/expectation within a reasonable error range. Accordingly, we call an object or phenomenon normal, if its attribute values lie in the normal range. In social media, we usually differentiate these two types of normal patterns in the context of network, geographic distribution and text content. With the definition of normal pattern, we can easily understand the anomaly as the outlier distribution of observing features/attributes. For example, Chae et al. [CCJ*15] (Figure 11d) firstly visualize geo-tagged Twitter messages and understand the normal distribution of social media users. Based on this, they propose a classification model to detect anomaly and visualize the abnormal trajectories, e.g. identifying the anomaly in the Boston Marathon. It is important to detect anomaly as it has a considerable impact on people and applications, especially in rumor detection of public opinion, sentiment anomaly and emerging crisis scenarios. The techniques are often used with normal pattern detection, visual comparison, and abnormal detection. For example, the work from Rohrdantz et al. [RHD*12] supports users to analyze sentiment patterns and reveal anomalies.

From the category matrix (Figure 10), we find some works falling in this category that address the spatial temporal distribution and movement trajectories. Surprisingly, there are few works explicitly emphasizing the detection of abnormal topics evolution. Most of works in topic analysis address detection of patterns. It reflects the challenges in detecting a subtle anomaly from large normal and noisy text data, which might be a research direction for further investigation. Besides these topics, there are research works focusing on investigating abnormal user behaviors based on the temporal distribution [CSL*16] and large changes of sentiments [ZGWZ14, SDB*15, RHD*12]. These techniques include a circular design with small multiples to highlight the anomaly through comparison and timeline chart with sentiment analysis to identify the abnormal sentiment in specified time ranges.

3.3.5. Predictive Analysis

Based on historical data, users can gain knowledge about general patterns, events, and anomalies. Furthermore, users, especially analysts, care about the trends and patterns in the future, which requires predictive analysis. Research in this category usually integrates the classification model with trained data to predict and to verify the prediction [LWM14, LKT*14, YJ15]. In the visual analytics pipeline of Yeon et al.'s work [YJ15], we can see the topic extraction process with general patterns and abnormal event detection is the basis for predictive analysis (Figure 11e). Then they predict and visualize event evolution patterns by combing the contextual similar cases occurring in the past. Predictive analysis combines tightly with predictive modeling and interactive visualization, so high-dimensional techniques such as parallel coordinates are used to help exploring the parameter space in the model [LWM14]. In general, there are few works addressing predictive analysis in visual analytics. Most of them address temporal evolution patterns [YJ15,LKT*14]. In geographic information, Wu et al. use geo-tagged data to predict movements and human mobility [WZSL14]. How to integrate interaction and visualization more tightly in predictive analysis is still a challenging and interesting area.

3.3.6. Situation Awareness

One important goal of visual analytics is to provide a user- and task-adaptable, guided representation that enables situation awareness [TC05]. Situation awareness integrates multiple above-mentioned visual analytics techniques to provide users enough information for decision making. SensePlace2 is one of the earliest situation awareness visual analytics tools based on the spatial temporal social media data [MJR*11]. It integrates spatial, temporal and textual information with dynamic monitoring, filtering and highlighting functions. In later time, ScatterBlogs and ScatterBlogs2 play important roles in situation awareness [TBK*12, BTH*13,TKE*15]. As illustrated in their pipeline [BTH*13] (Figure 11f), users can conduct live monitoring, apply classifiers and filters to analyze the stream. They also provide a full case study on how Twitter help analysts with situation awareness to save people's life in emergency scenarios [TKE*15]. TargetVue summarizes general user tweeting behaviors, identifies abnormal situations and provides situation awareness for the analysts [CSL*16]. Generally, we can find works targeting situation awareness to fully integrate multiple sources of data for analysis (Figure 3).

In short, social media visual analytics systems combine multiple entities, visualization, interaction and mining techniques to solve complex problems. In the next section, we show application scenarios and areas in social media visual analytics.

4. Systems and Applications

Solving real-world problems is one of the main goals of research in social media. We summarize the usage scenarios, case studies, and applications of selected visual analytics systems. Based on these, we summarize two general types of application scenarios, including the verification and development of social science theory and application as well as other domain-specific applications. Users' behaviors and the information diffusion process are the research targets in the general social science, including communication theory, migration analysis, information diffusion process, etc. Besides the general topics in social science, research results in social media visual analytics can apply to multiple disciplines. We summarize seven main types of application areas, including journalism, the emergency reaction for disasters, politics, finance, anti-terrorism, the crisis management, sports and entertainment as well as the tourism and urban planning applications (Figure 13). The used dataset includes Twitter, SinaWeibo, Google+, Flicker, etc. Among them, the most used data source is Twitter. We select several interesting cases for a detailed illustration.

4.1. Social Science Theory and Application

New visual analytics techniques targeting a large scale of social media data provide new perspectives for social science. There are

System	Application	Target User
Google+Ripples [VWH*13]	Socail Science Theory and Application	General Public
Li et al.2013 [LQC*13]	Socail Science Theory and Application	Analyst
Weibo KeyPlayer [YWL* 14]	Socail Science Theory and Application	Analyst
D-Map [CCW* 16]	Socail Science Theory and Application	Analyst
DemographicVis [DCE* 15]	Socail Science Theory and Application	Analyst
EvoRiver [SWL*14]	Socail Science Theory and Application	Analyst
TwitterScope [GHN13]	Social Science Theory and Application	Analyst
CompactMap [LWW* 13]	Social Science Theory and Application	Analyst
OpinionBlocks [HYZ*13]	Socail Science Theory and Application	General Public
PEARL [ZGWZ14]	Socail Science Theory and Application	General Public
WeiboEvents [RZW*14]	Journalism	General Public
CityBeat [XSX* 14]	Journalism	Analyst
Vox Civitas [DNK10]	Journalism	General Public
#FluxFlow [ZCW*14]	Disaster Management	Analyst
ScatterBlogs2 [BTH* 13]	Disaster Management	Analyst
Visual BackChannel [DGWC10]	Disaster Management	General Public
ideaFlow [WLC* 16]	Disaster Management, Politics	Analyst
Whisper [CLS* 12]	Disaster Management, Politics	General Public
Chae et al.2012 [CTB* 12]	Disaster Management, Crisis and Emergency Management	Analyst
Chae et al.2014 [CTJ* 14]	Disaster Management, Crisis and Emergency Management	Analyst
ScatterBlogs [TBK* 12]	Disaster Management, Crisis and Emergency Management	Analyst
ScatterBlogs2 [TKE* 15]	Disaster Management, Crisis and Emergency Management	Analyst
IS-I [WDM*12]	Disaster Management, Crisis and Emergency Management	Analyst
TargetVue [CSL* 16]	Crisis and Emergency Management	Analyst
GeoSocial [CCRS13]	Politics	Analyst
MutualRanker [LLZ* 16]	Politics	Analyst
ThemeCrowds [AGCH11]	Politics	Analyst
OpinionFlow [WLY*14]	Politics	Analyst
RoseRiver [CLWW14]	Politics	Analyst
SocialHelix [CLL* 14]	Politics	Analyst
LeadLine [DWS*12]	Polictis, Crisis and Emergency Management	General Public
Xu et al.2013 [XWW*13]	Polictis, Crisis and Emergency Management	Analyst
GraphFlow [CWL*14]	Polictis, Sports and Entertainment	Analyst
LeadLine [RWD14]	Finance	Analyst
Bollen et al.2011 [BMZ11]	Finance	Analyst
Wanner et al.2016[WJS*16]	Finance	General Public
NStreamAware [FS14]	Crisis and Emergency Management	Analyst
SensePlace2 [MJR* 11]	Crisis and Emergency Management	Analyst
GTAC [KWD* 13]	Crisis and Emergency Management	Analyst
Chae et al.2015 [CCJ* 15]	Crisis and Emergency Management	Analyst
Matisee [SDB*15]	Crisis and Emergency Management	Analyst
SentenTree [HWS17]	Sports and Entertainment	Analyst
Lu et al.2014 [LKT* 14]	Sports and Entertainment	Analyst
TwitInfo [MBB* 11]	Sports and Entertainment	Analyst
POI Pulse [MJG* 14]	Tourism and Urban Planning	General Public
Prieto et al. 2015 [PHE* 15]	Tourism and Urban Planning	Analyst
Chen et al. 2016 [CWLY16]	Tourism and Urban Planning	Analyst
FlowSampler [CMSVM14]	Tourism and Urban Planning	Analyst
TravelDiff [KSB* 16]	Tourism and Urban Planning	Analyst
SRS [BBD* 12]	Tourism and Urban Planning	Analyst

Figure 13: *Visual analytics systems collection, identifying the system name, application areas and target users. The color is identifying the main category of the work, which is coherent with the previous definition.*

many existing theories and hypotheses, which might be verified or improved in the setting of big data. For example, Sun et al. investigated the topic leader theory and an existing competition model in social science [SWL*14]. They use the large scale of Twitter data to verify the existing theory and refine the competition model by integrating the cooperative and competitive features of topics. They use the EvoRiver to evaluate and confirm the new model. Moreover, they investigate a case in topics including Gun, Government, Politics, Law and Order, Mental, etc. and observe how these topics collaborate and compete with each other. Besides verifying and improving the existing theories, visual analytics can help users identify new patterns in information diffusion of social media. Chen et al. design a map-based visual metaphor and investigate the egocentric information diffusion in the social network [CCW*16] (Figure 6e). By comparing 39 people with high impacts in Sina Weibo, they identify the social networks with new diffusion patterns and community behaviors, such as dual-center diffusion, strong center network, etc. The targeting users of the above systems are mainly analysts and social scientists. Besides, the social media platform is a content generation platform, which everyone can contribute to it. Opening visualization and visual analytics to the general public can gain more insights from their perspectives. Like Google+Ripples [VWH*13], they allow the Google+ users to explore the visualization with their own data. To better understand the users' usage, OpinionBlocks [HYZ*13] use crowd-sourcing to understand the opinions. Facing the general public, the visualization design should be intuitive and easy to understand. To achieve this, LeadLine construct a narrative visualization to clearly narrate the What, When, Who and Where attributes of the events [DWS*12] (Figure 6b).

For the social science theory and application, we find the common tasks include a feature extraction and anomaly detection. The motivation of them is to find patterns to verify existing theories or build up new hypotheses.

4.2. Domain-specific Applications

Because of the large engagement of people, news, and events, social media plays a more and more important role in the society. Analyzing social media data affects many disciplines.

4.2.1. Journalism

Social media is a new type of media, which changes the news spreading patterns. More journalists and major media open their accounts in the social media like Twitter. For journalists themselves, they need to understand well how the social media work and grasp the information diffusion patterns. However, the large amount of information makes it challenging to identify the useful information. To solve this problem, Diakopoulos et al. provide a comprehensive visualization system targeting journalists [DNK10]. The system includes a keyword search box, media and text visualization, video timeline, topic selector, Twitter timeline, sentiment timeline and dynamic keywords timeline (Figure 14a). With this tool, journalists and other people related to media can quickly identify Twitter keywords and understand the social events from multiple perspectives. They also show many cases, including discussions on Barack Obama's presidential debate. They infer the high support rate for Obama from the analysis of the events [DNYKS11]. Summarized from the above, visual monitoring and understanding the complex scenarios are important visualization and analytics goals [DNK10, DNYKS11]. Sharing the same goals, WeiboEvents provides a crowdsourcing event analysis platform by extracting features in the reposting network for the general public [RZW*14]. It enables users to crawl the social media message of interest and explore the visualization on-the-fly. Multiple users' findings can be aggregated to summarize the different event stages. With around 20,000+ usages, the WeiboEvents has helped many researchers and students in journalism to explore new ideas and verify theories. Besides exploring information diffusion, CityBeat [XSX*14] provides ambient monitoring of city operations, with which journalists can produce quick news with the spatial temporal events in social media.

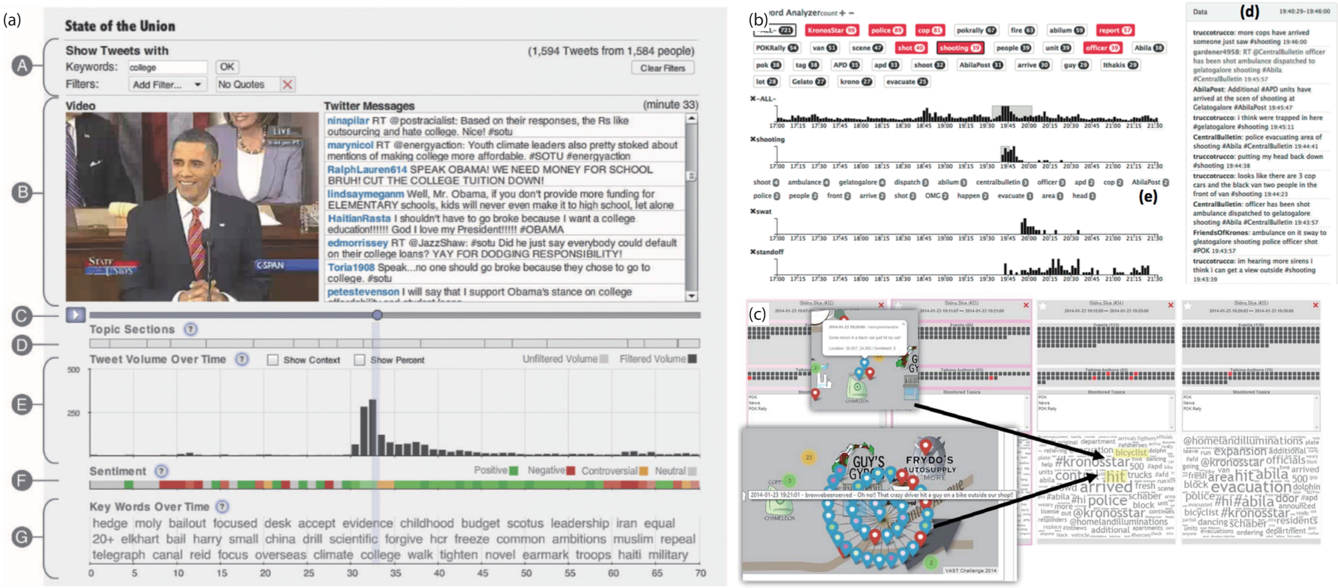

Figure 14: *Selected application systems of social media visual analytics for domain-specific applications. (a) Vox Civitas visualization system, providing journalists with an intuitive interface to explore the news and events [DNYKS11]. (b) Real-time keywords extraction from Twitter to support the situation awareness of emergency [LWC*14]. (c) Extracting diffusion network from real-time Twitter data, to enable users to identify the anomaly in crisis management [FS14].*

4.2.2. Emergency Reaction for Disasters

One advantage of social media is the quick information bursting and diffusion when an event happens. The time is critical when we face a disaster such as a flood, earthquake, etc. Messages in social media can help the emergency reaction to disaster in three ways. First, it conveys the disaster news and information to many people in a short time [DGWC10]. Visualization of such data helps users better understand the What, When, How and Where of the disaster [WDM*12]. Second, it helps to communicate the needs, the potential damage of affected regions for the related people and the helping volunteers. Third, it allows analysts to have the resources to analyze the scenarios and decide how to conduct better reacting actions. Visual analytics plays important roles in such scenarios. In the case study of Whisper [CLS*12], they found the messages burst in Japan, and then the surrounding countries including Philippines, Malaysia, etc. (Figure 5a). Because people in these countries worried about the effects of the hurricane caused by the earthquake. Besides the visual monitoring of such data, post-analysis of the detailed anomaly helps analysts better understand the disaster. Chae et al. investigated the Twitter messages during a long-term hurricane Sandy and a short term tornado [CTJ*14] (Figure 5c). By detecting the spatial temporal patterns and comparing with the abnormal topics, they found how the users react to various events were different. They found that the forecasting with evacuation order in social media worked effectively. When the real Sandy arrived in NYC, the users were prepared in most regions. But for the short bursting tornado, the users post many disaster POIs (Point of Interest) after the event in the city scale. These behaviors are clearly reflected in the visual analytics system. In short, emergency reactions are good examples showing how visual analytics help in event detection and anomaly detection. Usually special topics and spatial temporal distribution of messages will be reflected in emergency, thus the related techniques, such as wordles [AGCH11], map-based visualization [CLS*12] and timelines [CTJ*14] are used.

4.2.3. Politics

Politics are well discussed in social media. In one aspect, more politicians post messages to spread their ideas in the social media. In the other aspect, more people participate in discussing and sharing their opinions on political issues. There are many visual analytics system analyzing politics related cases [CCRS13, LLZ*16, WLY*14, CLWW14, AGCH11, SWL*14]. For example, EvoRiver represents the detail evolution of topics in the American Presidential Election [SWL*14]. They found "Government" and "Politics" were usually together. Because of the topic "International Issues", these two topics changed from competition to collaboration. They also found "Laws" and "Order" showed at the same time and they identified the key players discussing these topics. From the detail, they also found "Gun" management was one of the important issues in the debate. With such interactive visual analytics, analysts can identify the key players, issues and topic relations happened in the political process. In politics analysis scenarios, the complex relationships between users and topics along the time are critical features to tackle. Moreover, researchers care about the information diffusion and interactions among the topics. A series of newly designed river-based visualization techniques have been applied to visualize these features.

4.2.4. Finance

Data mining algorithms have been used to investigate how and when social media can affect the financial markets [Zhe15]. Bollen et al. investigate how the stock market is affected by the mood of users in Twitter [BMZ11]. However, in visual analytics applications, there is relative few works addressing this topic. One work is the application of LeadLine [DWS*12,RWD14]. They investigated the competitive advantages among the financial companies, especially the banks. With topic extraction and event timeline visualization, they identified several interesting events, such as the event at the revelation of JP Morgan Chase multi-billion dollar trading loss (the London Whale) [RWD14]. Wanner et al. [WJS*16] extract keywords in the financial news, detect news features, identify interesting financial time series intervals with financial statistic data and analyze the relationship between news and financial market. In short, topic extraction and event timeline summary are commonly used techniques, to identify the correlation of the social media messages/mood with the financial markets. Several predictive analysis scenarios [LWM14] are used, combined with the techniques such as parallel coordinates and other high dimensional techniques. More research can be addressed on the points since we see the large interests in combining the financial and social media data.

4.2.5. Anti-terrorism and Crisis Management

Social media messages are usually important data sources in anti-terrorism and crisis management. These applications require a heterogeneous data analysis, as well as specially designed interaction, data mining and visualization techniques for spatial-temporal and textual data. MacEachren et al. proposed SensePlace2, supporting the situation awareness of the spatial temporal distribution and text [MRJ*11]. Filtering in spatial and temporal scales is supported. In their case, analysts found self-organized fund-raising behaviors in the Haiti Earthquake through multiple combinations of filters. The IEEE Visual Analytics Challenge (IEEE VAST Challenge) has provided several challenges in crisis management with social media data [CGW14]. For example, in 2014, they required participants to analyze a kid-nap event and several crisis issues in Mini-Challenge 3. Among the award winners, Liu et al. provide a visual analytics system combining spatial-temporal, textual and keywords, which enables the collaborative analysis for streaming data. It enables users to identify the overall event stages and the detailed topics dynamically [LWC*14] (Figure 14b). Fischer et al. provide a card-based visual analytics system. With the flowed card information, users can identify the suspicious people and their behaviors in the graph context [FS14](Figure 14c).

4.2.6. Sports and Entertainment

People are eager to discuss sport games, concert, films, etc. in social media. Moreover, there are live reports of several events in the famous sport games in social media, which people can actively participate in. Visual analytics can help users better understand the game stages, special events or characters in the entertainment process [LKT*14, HWS17, MBB*11, CWL*14]. The existing examples are about the text and topic analysis in the context of sport games as event detection and analysis. Marcus et al. propose TwitInfo for event detection and sentiment analysis [MBB*11]. They test a case in Twitter data discussing a sport game (Figure 12b). Their event detection component correctly detects and visualizes the sub-event peaks, such as game start, goal, halftime, etc. Being also applied in the sports, SentenTree focuses on the keywords relationship [HWS17]. They construct a tree of keywords with a node-link diagram to give users clues of co-occurrence of these words. They can describe the scenarios with the semantics. For example, they find words such as "Neymar", "score", "penalty" appeared in order, which can indicate the order of the happening events. Besides the sports, Lu et al. apply visual analytics of social media in box office forecasting [LWM14] (Figure 12c). They classify people's sentiment and keywords to conduct the prediction.

4.2.7. Tourism and Urban Planning

Geo-tagged social media cover a large number of people and has wide geographic ranges. There are visual analytics systems targeting the general public [MJG*14] and analysts or researchers [CYW*16, CMSVM14, PHE*15, BBD*12, KSB*16]. POI Pulse [MJG*14] shows the distributed density map within the city and small multiples of different POIs. General people can easily understand different activities in each POI derived from social media. Best et al. also provide a topic stream visualization in Seattle city, for the government to monitor the public opinions [BBD*12]. Prieto et al. use event distribution visualization to support urban planning [PHE*15]. Besides the event distribution, trip making and tourism analysis are important applications of geo-tagged social media visual analytics. Chen et al. analyze the social movement patterns of inter-city footprint [CWLY16]. They also identify interesting tourists patterns and verify them from the famous trip suggesting websites (Figure 12a). In short, the tasks include to derive the movement patterns for the suggestion of improving tourist route planning or land usage patterns to improve the decision on urban planning.

4.3. Public and Commercial Social Media Analytics Tools

The development of public and commercial tools in social media visualization and visual analytics increases. These kinds of tools can monitor, analyze and manage social media information statistics and the impact of social brands. They generally focus on gathering data from a specific or multiple websites, visualizing important messages, and generating reports. Besides, they also give suggestions to help users get a picture of their impacts, and guide them to achieve more impacts. We have collected 27 popular social media visualization and analytics tools (Figure 15). Most of these tools get data from several popular social medias, such as Facebook, Twitter, etc. Some of them focus on a specific social media platform, such as Iconosquare [Ico24] and SocioViz [Soc24a]. Besides, we also list some generally useful tools, such as Gephi [Gep24] and Google Analytics [Goo24].

Based on the main functions of social media analytics tools, we divide them into three categories, including analytics, monitoring, and management. Social media analytics focus on providing analytical reasoning from data, especially analyzing the topic distribution, keywords trends to gain understanding of the social behaviors. Representative tools include

Name	Type	Main Functions	Data Type/Souce	Free
Crowdbooster [Cro]	Analytics	Follower Evaluation; Collaboration; Sharing	Twitter, Facebook	NO
Cytoscape [Cyt]	Analytics	Network Integration, Visualization and Analysis	Network	YES
Gephi [Gep]	Analytics	Large Network Analysis	Network	YES
Google Analytics [Anab]	Analytics	Visiting Exploration	/	YES
Klout [Klo]	Analytics	Great Content Creation and Sharing	Bing, Facebook, Foursquare, Google+, Instagram, LinkedIn, Twitter, Youtube, Wikipedia	NO
Netlytic [Net]	Analytics	Social Network Exploration; Popular Topics Discovering	RSS Feeds, Facebook, Twitter, Youtube, Instagram, text/csv file	YES
SocioViz [Socd]	Analytics	Key Influencers Identification; Text Analysis	Twitter	YES
SocNetV [Soce]	Analytics	Network Exploration with Various Layout Models	GraphML, Pajek, UCINET, GraphViz, Adjacency, EdgeList	YES
ZhiweiData [Zhi]	Analytics	Massive Data Visualization; Social Events Monitoring	/	YES
AgoraPulse [Ago]	Management	Content Integration; Report Export	Facebook, Twitter, Instagram, LinkedIn, Google+	NO
Buffer for Business [fB]	Management	Traffic Driving; Fan Engagement	/	NO
BuzzSumo [Buz]	Management	Topic-wised Content Analysis	/	NO
Crimson Hexagon [Hex]	Management	Strategic Business Insight; Topic-wised Content Analysis	Twitter, Facebook, Instagram, Weibo, Blogs, Forums	NO
HootSuite [Hoo]	Management	Influencers Identification	Facebook, Twitter, Instagram, Youtube, LinkedIn, Google+	NO
HubSpot [Hub]	Management	Inbox marketing	Inbound, Youtube, Twitter, LinkedIn, SlideShare, Pinterest, Readthink, iTunes	NO
Iconosquare [Ico]	Management	Instagram Analysis	Instagram	NO
Synthesio [Syn]	Management	Customized Dashboards; Monitoring in Real-time	Website	NO
Tailwind [Tai]	Management	Conversations Monitoring	Pinterest, Instagram	NO
Zoho Social [Socc]	Management	Right Audience Identification; Engagement	Twitrer, Facebook, LinkedIn, Google+, Instagram	NO
Brand24 [Bra]	Monitor	Real-time Reaction; Sales Opportunities Detection	Millions of sources	NO
Brandwatch Analytics [Anaa]	Monitor	Brand Insight Extraction	80 million online sources	NO
Digimind Intelligence [Int]	Monitor	Real-time Text Mining	Twitter, Pinterest, Instagram and all major social networks	NO
Digimind Social [Soca]	Monitor	Conversations Monitoring; Online Reputation Analysis	Unlimited social media & web	NO
Keyhole [Key]	Monitor	Influencers Identification	Twitter, Instagram	NO
Sprout Social [Socb]	Monitor	Various Social Medias Integration and Monitoring	Twitter, Facebook, Messager, Google+, Instagram	NO
Sysomos [Sys]	Monitor	Influencers & Trends; Identification & Monitoring	/	NO
Talkwalker [Tal]	Monitor	Image Recognization and Analysis	150 million websites	NO

Figure 15: *Selected public and commercial tools for social media visualization and visual analytics. "/" represents missing information in their websites or products descriptions. We select 27 representative applications and list their system names, types, functions, data and whether it is free.*

Crowdbooster [Cro24], Cytoscape [Cyt24], Klout [Klo24], Netlytic [Net24], SocNetV [Soc24b], ZhiweiData [Zhi24], etc. For example, ZhiweiData [Zhi24] is a visualization system targeting Sina Weibo. It crawls the data with customized keywords and provides a timeline highlighting the important stages of the events. It also shows the people's attitude, robots percentages, key players etc. to help users analyze the events. Most of them use node-link diagram for social network exploration, such as Cytoscape [Cyt24], Netlytic [Net24] and SocioViz [Soc24a]. Other techniques, such as river-like visualization and map-based visualization, are also used in the tool like Netlytic [Net24].

Social media monitoring, also known as social listening, is to track, gather and analyze online conversations in social media about their brands or related topics and so on. With the data, the analysts try to identify and assess reputation and influence of certain individuals or groups. Representative tools include Brand24 [Bra24a], Brandwatch Analytics [Bra24b], Digimind Intelligence [Dig24a], Digimind Social [Dig24b], Keyhole [Key24], Sprout Social [Spr24], Sysomos [Sys24], Talkwalker [Tal24], etc. Various visualization techniques, such as map-based visualization, timeline chart, river-like visualization, wordle, are used in these kinds of tools, including Keyhole [Key24] and Talkwalker [Tal24].

Social media management focuses on helping users gain an insight of their influence, behavior of their customers and competitors. It also analyzes the data and generates reports to help users gain better influence. Representative tools include AgoraPulse [Ago24], Buffer for Business [Buf24], BuzzSumo [Buz24], Crimson Hexagon [Cri24], HootSuite [Hoo24], HubSpot [Hub24], Synthesio [Syn24], Tailwind [Tai24], Zoho Social [Zoh24], etc. This kind of tools always would combine various visualization techniques to visualize heterogeneous data. For example, Iconosquare [Ico24] uses a timeline chart to show statistical information and the tool Synthesio [Syn24] uses a map-based visualization, wordle and timeline chart to show messages for users to gain information and manage their own information.

5. Observations and Discussions

In this survey work, we first discuss the data entities and corresponding visualizations in social media. Based on three categories with nine subcategories, we analyze the related research problems and visualization techniques for each. One step further, we analyze how different techniques are combined together, for serving different types of visual analytics goals. Based on these analytics approaches, we select representative visual analytics systems as well as commercial platforms, and then outline multiple application domains for such works. In the process, we make multiple interesting observations and summaries.

5.1. Features and Limitations of Social Media Data

Social media data content can contain multimedia, such as image and video. The multimedia is used to detect semantically meaningful topics and analyze the evolution of events in social media [QZXS16, PJZ*15, CYLH15, NFL*14]. It is an important topic in social media analytics, but not addressed fully with visual analytics yet. We collect and discuss some of the advanced techniques, to shed insight on future research. Researchers integrate the multimedia with feature extraction and event detection. For example, mmETM [QZXS16] can effectively model multimedia data, such as long text with related images, to extract topics of significant events. The model is used in social event tracking and evolutionary trends analysis. Niu et al. [NFL*14] develop a multi-source-driven asynchronous diffusion model (MADM) to study the video-sharing propagation in social media. More details can be found in Wu et al.'s survey [WCG*16].

Though social media data provide fruitful sources for analysis in a wide range of domains [KFS*12, KH10], there are several limitations and issues we need to pay attention to when we use social media data. First, there is bias in representing people's behaviors when we use social media data. Young people, and people who have access to mobile phones and microblog services have more chances to post social media messages. For geo-tagged analysis, there is around 3%-5% people posting messages with geo-tagged information in Sina Weibo [CYW*16]. Moreover, the data we get are usually samples from one social media platform (e.g. Twitter). Considering this, researchers should pay attention to the data distribution, research scope and the bias issues when conducting research in social media. That is to say, when we generalize from social media data to social science, we need further take consideration with risk and uncertainty in mind. Second, the trustiness of the social media message should be considered. Geo-tagged social media data is sometimes not trustable because users might fake their GPS. The content needs further verification to be rumors or not should also be considered when we are using social media messages. Some robots or scripts also intentionally post messages or trigger events. Lastly, the privacy issue is always the problem, which is not covered in this survey. There are many researches covering the privacy issues [MLC*13, Geo06, KM06]. In short, these are the both challenges and opportunities to use visual analytics to solve problems with social media data.

5.2. Network Researches Evolution in Social Media

We can find an interesting evolutionary patterns of social media visualization, especially in social network visualization. In the beginning, researchers focused on social network visualization, applying the existing techniques to analyze the follower networks in social media [HB05] (Figure 4a). In the next stage, more and more social events were discussed on Twitter, etc. Researchers paid attention to the information diffusion process visualization and analysis [RZW*14, CCW*16] (Figure 4d, e). Research in follower network visualization became less and the commercial tools analyzing such networks increased. We can see that the research in this area becomes mature and transfers into commercial products.

We also observe the trend that analyzing the information diffusion and reposting behaviors with semantic information [CCW*16, WLC*16] is still an active theme recently. Different perspectives, including ego-centric analysis [CCW*16], community-level analysis [SWL*14] and topic-evolution analysis [WLY*14], etc, have been investigated. In-depth social event analysis, regarding the diffusion network, will obtain further attention. Moreover, the diffusion network is correlated with the follower network in social media. People are likely to repost friends' messages and also tend to follow the people that share similar opinions. Analyzing how information diffuses among the follower network will enable users to better understand the information diffusion process. However, we haven't seen too many papers addressing this topic.

Compared with other networks, e.g. citation networks [MGF12, vEW14], social network in social media are different in both data characteristics and analytical requirement. A citation network is acyclic and with time information. Both citation networks and social networks in social media are dynamic and face the challenges in dynamic network visual analysis, such as analyzing the evolution of dynamic graph [BBDW14]. For the networks in social media, the diffusion network has the same acyclic constraints (multi-tree structure). While the follower network and reposting network do not have such constraints. The other difference is that the social network in social media has strong temporal features, e.g. bursting characteristics, fast-dying features, a large amount of users for significant events, multi-variate entities, etc. The analytical tasks can by either exploring the intrinsic patterns or help other research with social media data.

5.3. New perspectives of Semantic Movement Analysis

Visualizing geo-tagged messages on the map is a common approach for users to identify the keywords distribution [CTJ*14] (Figure 5c). One step further, researchers investigate how to understand the event distribution with advanced classification and interactions [TBK*12, BTH*13, CCJ*15]. One trend we observe is from Krueger [KTE15]. They use the geo-tagged social media data to enrich other movement datasets, such as car trajectories, helping users to better understand the semantic of the movement. Besides, analyzing sparsely sampled trajectories, defined as episodic movement [AAS*12], is also a potential topic. It is fruitful to investigate aggregated movement patterns by promptly dealing with uncertainties. But the uncertainties problems are not fully solved by the current solutions, e.g., especially the inner-city trajectories analysis [CYW*16] (Figure 12a). From the application perspective, e.g. in tourism application, there are large demands on analyzing people's movement based on social media, because analysts from toursim domain not only want to know the moving paths of the tourists but also want to gain ideas of their thoughts and comments on the journey. Visual analytics is helpful to improve tourists' visiting paths and the settings of the places of interest.

5.4. Text: Diverse Visual Design and Analytical Reasoning

The river-based visual metaphor is a good design choice to visualize dynamic text, topics, and sentiment. A series of interesting works have been conducted with river-based visual metaphor in dynamic topic evolution [XWW*13, WLY*14, SWL*14], hierarchical topic analysis [DYW*13, CLWW14] and semantic analysis [ZGWZ14, SDB*15], etc. (Figure 6). Though intuitive, the river-based visual metaphor has its shortcomings. One important issue is that it restricts other information, e.g. text, relationship, geo-information, etc., into one dimension because the other dimension is time. Circular visual design [CLS*12] (Figure 5a) and map-based projection (Figure 4e, f) might partially solve the problem [GHN13, LWW*13, CCW*16, CSL*16]. Liu et al. project all the tweets onto a map based on the topics [LWW*13]. It makes good use of space to illustrate the correlations of topics and dynamically update layout based on time. However, with these approaches, the time dimension is implicit and only animation might not be enough to present the temporal variation. From this point, a variety of visual design can be combined to further represent the complex dynamic features of textual and relational information in the future. Moreover, real-time streaming data sources might be common data source recently [BTH*13]. Thus, designing a scalable and progressive visual analytics system targeting this feature is also important. It requires conquering the difficulties in real time computing, visual design and dynamic interactions.

We summarize three levels to analyze the text content. However, the gap between keywords/topics and the actual contents of messages is still large. Existing works try to analyze the content in two ways. First, the summarized keywords and topics provide abstracted semantic information derived from the contents. It is defined as important information. Second, researchers provide summarized keywords and topics as filters and users can select related contents to investigate the raw data. However, there are still gaps to fully understand all the content information, which is restricted by current visual analytics and NLP (Natural Language Process) techniques. Currently, we still do not have sufficient control to investigate all the message contents. Future visual analytics direction can be providing explorable methods to navigate the content in multiple perspective, rather than only with keywords/topics with high frequency. Currently, the insight from social media in three categories has been achieved by case studies in many fields. By summarizing these cases, we find the current techniques focus on what, when, where and who. However, the most important "Why", which indicates the reasons and behaviors led to the results, is usually not fully covered. For example, in event detection, many systems are proposed to detect the event, but not easily identify the exact behaviors that push forward the event evolution. From this perspective, the works in the future might further conduct the analytical reasoning with the social media data.

5.5. Heterogeneous Data Visual Analysis

To summarize the heterogeneous data visual analysis in social media, we can summarize three levels.

- **Multiple attributes within the same users** It involves the combination of multiple entities including network, spatial temporal movement, and topic or sentiment distribution of the same users. It provides multiple perspectives for analyzing people's behaviors in social media.
- **Enriching the social media with additional information** To better understand the social behaviors of users, we might need additional information. For example, in trajectory analysis of social media, we might need to derive the POI information to further identify the semantics of the movement.
- **Social Media+: Combining social media in multi-discipline analysis** Social media involves the social activities of users. It has connections with many disciplines. More than the examples we illustrate in the application survey, we believe it can be used with multiple application areas, especially large ranges of social science and application. Social media users act as the sensors all over the world capturing and receiving spatial, temporal and textual information. Knowledge derived from social media can help better decision making, which has been proved in many commercial tools. Finally, we envision combining both the cyberspace and physical space within a uniform analytical environment.

5.6. Evaluation of Social Media Visual Analytics

Currently, almost all of the research papers in social media analysis verify their contributions and claims with case studies. In current stages, many new problems are proposed and solved in a case-by-case manner. There is still no general evaluation rules and guidelines for social media visual analytics yet. By summarizing these papers, we try to identify several key features that a novel social media visual analytics tools should provide. First, targeting users of the visual analytics platform should be clearly defined. It is different between public users and the analysts. Second, how much scalable of the proposed method, to which extent of data amount it supports might be evaluated. Third, users have the desired patterns, features or events found by the proposed methods. A theoretical analysis and summaries of such desired patterns are encouraged to provide. It helps to judge the scope, pros, and cons of the proposed methods.

6. Conclusion

In this survey, we first identify the features of social media data and summarize the needs and impact on analyzing such data. Among all the analysis, visual analytics is an important way to derive insight with visualization, interaction and data mining techniques. We summarize the general research pipeline of social media visual analytics and identify nine categories of entities and the corresponding visualization techniques with details. Afterward, we discuss the combination of multiple entities with visual analytics serving for six types of goals. We also derive insights from applications in multiple disciplines as well as commercial products. By analyzing these paper, we frame a taxonomy of social media visual analytics and derive more guidelines for possible future trends.

Acknowledgement

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

References

[AAF*13] ANDRIENKO G. L., ANDRIENKO N. V., FUCHS G., RAIMOND A. O., SYMANZIK J., ZIEMLICKI C.: Extracting semantics of individual places from movement data by analyzing temporal patterns of visits. In *Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place* (2013), COMP '13, ACM, pp. 9:9–9:16. 7

[AAS*12] ANDRIENKO N., ANDRIENKO G., STANGE H., LIEBIG T., HECKER D.: Visual analytics for understanding spatial situations from episodic movement data. *KI-Künstliche Intelligenz 26*, 3 (2012), 241–251. 7, 20

[AGCH11] Archambault D., Greene D., Cunningham P., Hurley N.: Themecrowds: Multiresolution summaries of twitter usage. In *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents* (2011), SMUC '11, ACM, pp. 77–84. 2, 8, 9, 17

[Ago24] Agorapulse, [Online; accessed 2017-01-24]. http://www.agorapulse.com/. 19

[BBD*12] Best D. M., Bruce J. R., Dowson S. T., Love O. J., McGrath L. R.: *Web-based visual analytics for social media*. Tech. rep., Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2012. 8, 12, 18

[BBDW14] Beck F., Burch M., Diehl S., Weiskopf D.: The State of the Art in Visualizing Dynamic Graphs. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. 3, 20

[BBDW16] Beck F., Burch M., Diehl S., Weiskopf D.: A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum* (2016). 1

[BCD*10] Bezerianos A., Chevalier F., Dragicevic P., Elmqvist N., Fekete J.: Graphdice: A system for exploring multivariate social networks. *Computer Graphics Forum 29*, 3 (2010), 863–872. 4, 11

[BMZ11] Bollen J., Mao H., Zeng X.: Twitter mood predicts the stock market. *Journal of computational science 2*, 1 (2011), 1–8. 18

[BN11] Brandes U., Nick B.: Asymmetric relations in longitudinal social networks. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 2283–2290. 4, 5

[Bra24a] Brand24, [Online; accessed 2017-01-24]. http://brand24.com/. 19

[Bra24b] Brandwatch analytics, [Online; accessed 2017-01-24]. http://www.brandwatch.com/brandwatch-analytics/. 19

[BT14] Batrinca B., Treleaven P. C.: Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY 30*, 1 (2014), 89–116. 1

[BTH*13] Bosch H., Thom D., Heimerl F., Püttmann E., Koch S., Krüger R., Wörner M., Ertl T.: Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering. *IEEE Transactions on Visualization and Computer Graphics 19* (2013), 2022–2031. 2, 7, 8, 12, 13, 14, 15, 20, 21

[Buf24] Buffer for business, [Online; accessed 2017-01-24]. http://buffer.com/business/. 19

[Buz24] Buzzsumo, [Online; accessed 2017-01-24]. http://buzzsumo.com/. 19

[CC16] Cao N., Cui W.: Introduction to text visualization. *Atlantis briefs in artificial intelligence 1* (2016). 8

[CCJ*15] Chae J., Cui Y., Jang Y., Wang G., Malik A., Ebert D. S.: Trajectory-based visual analytics for anomalous human movement analysis using social media. *The Eurographics Association* (2015). 7, 11, 13, 14, 15, 20

[CCRS13] Croitoru A., Crooks A., Radzikowski J., Stefanidis A.: Geosocial gauge: A system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science 27* (2013), 2483–2508. 6, 7, 17

[CCW*16] Chen S., Chen S., Wang Z., Liang J., Yuan X., Cao N., Wu Y.: D-map: Visual analysis of ego-centric information diffusion patterns in social media. In *Proc. of IEEE Visual Analytics Science and Technology* (2016). 5, 6, 11, 12, 13, 16, 20

[CGW14] Cook K. A., Grinstein G. G., Whiting M. A.: The VAST challenge: history, scope, and outcomes: An introduction to the special issue. *Information Visualization 13*, 4 (2014), 301–312. 18

[CLS*12] Cao N., Lin Y., Sun X., Lazer D., Liu S., Qu H.: Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2649–2658. 2, 6, 7, 11, 12, 13, 17, 20

[CLWW14] Cui W., Liu S., Wu Z., Wei H.: How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2281–2290. 8, 13, 17, 20

[CMSVM14] Chua A., Marcheggiani E., Servillo L. A., Vande Moere A.: Flowsampler: Visual analysis of urban flows in geolocated social media data. In *6th International Conference on Social Informatics (workshops)* (2014). 7, 8, 18

[Cri24] Crimson hexagon, [Online; accessed 2017-01-24]. http://www.crimsonhexagon.com/. 19

[Cro24] Crowdbooster, [Online; accessed 2017-01-24]. http://crowdbooster.com/. 19

[CSL*16] Cao N., Shi C., Lin S., Lu J., Lin Y.-R., Lin C.-Y.: Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 280–289. 5, 6, 15, 20

[CTB*12] Chae J., Thom D., Bosch H., Jang Y., Maciejewski R., Ebert D. S., Ertl T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2012), pp. 143–152. 6, 14

[CTJ*14] Chae J., Thom D., Jang Y., Kim S., Ertl T., Ebert D. S.: Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics 38* (2014), 51 – 60. 6, 7, 13, 14, 17, 20

[CWL*14] Cui W., Wang X., Liu S., Riche N. H., Madhyastha T. M., Ma K.-L., Guo B.: Let it flow: A static method for exploring dynamic graphs. In *Visualization Symposium (PacificVis), IEEE PacificVis* (2014), pp. 121–128. 5, 6, 18

[CWLY16] Chen S., Wang Z., Liang J., Yuan X.: Weibo footprint: A web-based visualization system to analyzing spatial-temporal movement of geo-tagged social media users. In *Visualization Symposium (PacificVis), IEEE PacificVis Poster* (2016). 11, 13, 18

[CYLH15] Cai H., Yang Y., Li X., Huang Z.: What are popular: exploring twitter features for event detection, tracking and visualization. In *Proceedings of the 23rd ACM international conference on Multimedia* (2015), ACM, pp. 89–98. 19

[Cyt24] Cytoscape, [Online; accessed 2017-01-24]. http://www.cytoscape.org/. 19

[CYW*16] Chen S., Yuan X., Wang Z., Guo C., Liang J., Wang Z., Zhang X. L., Zhang J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan 2016), 270–279. 2, 6, 7, 8, 11, 12, 13, 14, 18, 20

[CZH*09] Chi Y., Zhu S., Hino K., Gong Y., Zhang Y.: iolap: a framework for analyzing the internet, social networks, and other networked data. *IEEE Transactions on Multimedia 11*, 3 (2009), 372–382. 4

[DCE*15] Dou W., Cho I., ElTayeby O., Choo J., Wang X., Ribarsky W.: Demographicvis: Analyzing demographic information based on user generated content. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2015), pp. 57–64. 4

[DGWC10] Dörk M., Gruen D. M., Williamson C., Carpendale M. S. T.: A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1129–1138. 8, 12, 17

[Dig24a] Digimind intelligence, [Online; accessed 2017-01-24]. http://www.digimind.com/features-intelligence/. 19

[Dig24b] Digimind social, [Online; accessed 2017-01-24]. http://www.digimind.com/features-social/. 19

[DNK10] Diakopoulos N., Naaman M., Kivran-Swaine F.: Diamonds in the rough: Social media visual analytics for journalistic inquiry.

In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2010), pp. 115–122. 16

[DNYKS11] DIAKOPOULOS N., NAAMAN M., YAZDANI T., KIVRAN-SWAINE F.: Social media visual analytics for events. In *Social Media Modeling and Computing*. Springer, 2011, pp. 189–209. 16, 17

[DWS*12] DOU W., WANG X., SKAU D., RIBARSKY W., ZHOU M. X.: Leadline: Interactive visual analysis of text data through event identification and exploration. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2012), pp. 93–102. 2, 8, 9, 14, 16, 18

[DYW*13] DOU W., YU L., WANG X., MA Z., RIBARSKY W.: Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics 19* (2013), 2002–2011. 8, 11, 12, 14, 20

[Eve12] Event, [Online; accessed 2017-04-12]. http://en.wikipedia.org/wiki/Event. 14

[FS14] FISCHER F., STOFFEL F.: NStreamAware: Real-Time Visual Analytics for Data Streams (VAST Challenge 2014). In *IEEE Conference on Visual Analytics Science and Technology (VAST Challenge 2014 MC3)* (2014). Award for Outstanding Comprehensive Submission. 12, 17, 18

[Geo06] GEORGE A.: Living online: The end of privacy. *New Scientist 2569* (2006), 1–50. 20

[Gep24] Gephi, [Online; accessed 2017-01-24]. http://gephi.org/. 18

[GFC05] GHONIEM M., FEKETE J.-D., CASTAGLIOLA P.: On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. *Information Visualization 4*, 2 (2005), 114–135. 3

[GHFZ13] GUILLE A., HACID H., FAVRE C., ZIGHED D. A.: Information diffusion in online social networks: a survey. *SIGMOD Record 42*, 2 (2013), 17–28. 1

[GHN13] GANSNER E. R., HU Y., NORTH S. C.: Interactive visualization of streaming text data with dynamic maps. *J. Graph Algorithms Appl. 17*, 4 (2013), 515–540. 20

[Goo24] Google analytics, [Online; accessed 2017-01-24]. http://www.google.com/analytics/. 18

[HB05] HEER J., BOYD D.: Vizster: Visualizing online social networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on* (2005), IEEE, pp. 32–39. 3, 5, 6, 20

[HF06] HENRY N., FEKETE J.-D.: Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics 12* (2006), 677 –684. 3

[HF07] HENRY N., FEKETE J.-D.: Matlink: Enhanced matrix visualization for analyzing social networks. In *Proc. International Conference on Human-Computer Interaction* (2007), vol. 4663, pp. 288–302. 3, 6, 11

[HFM07] HENRY N., FEKETE J.-D., MCGUFFIN M. J.: Nodetrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1302–1309. 3, 6, 11

[HHN00] HAVRE S., HETZLER B., NOWELL L.: Themeriver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings* (2000), pp. 115–123. 8

[Hoo24] Hootsuite, [Online; accessed 2017-01-24]. http://hootsuite.com/. 19

[HTA*15] HUANG D., TORY M., ASENIERO B. A., BARTRAM L., BATEMAN S., CARPENDALE S., TANG A., WOODBURY R.: Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics 21* (2015), 420–433. 1

[Hub24] Hubspot, [Online; accessed 2017-01-24]. http://www.hubspot.com/. 19

[HWS17] HU M., WONGSUPHASAWAT K., STASKO J.: Visualizing social media content with sententree. *IEEE Transactions on Visualization and Computer Graphics PP*, 99 (2017), 1–1. 8, 18

[HYZ*13] HU M., YANG H., ZHOU M. X., GOU L., LI Y., HABER E.: *Human-Computer Interaction – INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part II*. Springer Berlin Heidelberg, 2013, ch. OpinionBlocks: A Crowd-Powered, Self-improving Interactive Visual Analytic System for Understanding Opinion Text, pp. 116–134. 9, 16

[Ico24] Iconosquare, [Online; accessed 2017-01-24]. http://pro.iconosquare.com/. 18, 19

[KAF*08] KEIM D., ANDRIENKO G., FEKETE J.-D., GÖRG C., KOHLHAMMER J., MELANÇON G.: Information visualization. Springer-Verlag, 2008, ch. Visual Analytics: Definition, Process, and Challenges, pp. 154–175. 12

[KEC06] KELLER R., ECKERT C. M., CLARKSON P. J.: Matrices or node-link diagrams: which visual representation is better for visualising connectivity models? *Information Visualization 5*, 1 (2006), 62–76. 3

[Key24] Keyhole, [Online; accessed 2017-01-24]. http://keyhole.co/. 19

[KFS*12] KAVANAUGH A. L., FOX E. A., SHEETZ S. D., YANG S., LI L. T., SHOEMAKER D. J., NATSEV A., XIE L.: Social media use by government: From the routine to the critical. *Government Information Quarterly 29*, 4 (2012), 480–491. 20

[KH10] KAPLAN A. M., HAENLEIN M.: Users of the world, unite! the challenges and opportunities of social media. *Business horizons 53*, 1 (2010), 59–68. 20

[KK15] KUCHER K., KERREN A.: Text visualization techniques: Taxonomy, visual survey, and community insights. In *Visualization Symposium (PacificVis), IEEE PacificVis* (2015), pp. 117–121. 1

[KKRS13] KEIM D., KRSTAJIC M., ROHRDANTZ C., SCHRECK T.: Real-time visual analytics for text streams. *Computer 46*, 7 (July 2013), 47–55. 1

[Klo24] Klout, [Online; accessed 2017-01-24]. http://klout.com/. 19

[KM06] KORNBLUM J., MARKLEIN M. B.: What you say online could haunt you. *USA Today 9* (2006), 1A. 20

[KSB*16] KRUEGER R., SUN G., BECK F., LIANG R., ERTL T.: Traveldiff: Visual comparison analytics for massive movement patterns derived from twitter. In *Visualization Symposium (PacificVis), IEEE PacificVis* (April 2016), pp. 176–183. 7, 11, 13, 18

[KSBK*15] KUCHER K., SCHAMP-BJEREDE T., KERREN A., PARADIS C., SAHLGREN M.: Visual analysis of online social media to open up the investigation of stance phenomena. *Information Visualization* (2015). 9

[KTE15] KRUEGER R., THOM D., ERTL T.: Semantic enrichment of movement behavior with foursquare: A visual analytics approach. *IEEE Transactions on Visualization and Computer Graphics* (2015), 903–915. 7, 20

[KWD*13] KRAFT T., WANG D. X., DELAWDER J., DOU W., YU L., RIBARSKY W.: Less after-the-fact: Investigative visual analysis of events from streaming twitter. In *Large-Scale Data Analysis and Visualization (LDAV), 2013 IEEE Symposium on* (2013), pp. 95–103. 6, 7, 13, 14

[LKT*14] LU Y., KRÜGER R., THOM D., WANG F., KOCH S., ERTL T., MACIEJEWSKI R.: Integrating predictive analytics and social media. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 193–202. 15, 18

[LLZ*16] LIU M., LIU S., ZHU X., LIAO Q., WEI F., PAN S.: An uncertainty-aware approach for exploratory microblog retrieval. *IEEE Transactions on Visualization and Computer Graphics 22* (2016), 250–259. 8, 17

[LQC*13] LI Q., QU H., CHEN L., WANG R., YONG J., SI D.: Visual analysis of retweeting propagation network in a microblogging platform. In *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction* (2013), VINCI '13, ACM, pp. 44–53. 4, 14

[LSKG14] LIU Y., SUI Z., KANG C., GAO Y.: Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE 9*, 1 (2014), e86026. 7, 11

[LWC*14] LIU Z., WANG Z., CHEN S., WANG Z., MIAO Z., YUAN X.: A platform for collaborative visual analysis on streaming messages. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 375–376. 17, 18

[LWM14] LU Y., WANG F., MACIEJEWSKI R.: Business intelligence from social media: A study from the vast box office challenge. *IEEE Computer Graphics and Applications 34* (2014), 58–69. 14, 15, 18

[LWW*13] LIU S., WU Y., WEI E., LIU M., LIU Y.: Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE InfoVis 2013 19*, 12 (2013), 2436–2445. 20, 21

[LWYZ15] LIU Y., WANG C., YE P., ZHANG K.: Analysis of microblog diffusion using a dynamic fluid model. *Journal of Visualization 18*, 2 (2015), 201–219. 4

[LXG*16] LIU X., XU A., GOU L., LIU H., AKKIRAJU R., SHEN H.-W.: Socialbrands: Visual analysis of public perceptions of brands on social media. In *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on* (2016), IEEE, pp. 71–80. 9

[LYW*16] LIU S., YIN J., WANG X., CUI W., CAO K., PEI J.: Online visual analytics of text streams. *IEEE transactions on visualization and computer graphics 22*, 11 (2016), 2451–2466. 8

[MBB*11] MARCUS A., BERNSTEIN M. S., BADAR O., KARGER D. R., MADDEN S., MILLER R. C.: Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), CHI '11, ACM, pp. 227–236. 14, 18

[MGF12] MATEJKA J., GROSSMAN T., FITZMAURICE G.: Citeology: visualizing paper genealogy. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (2012), ACM, pp. 181–190. 20

[MGMZ14] MUELDER C., GOU L., MA K.-L., ZHOU M. X.: Multivariate social network visual analytics. In *Multivariate Network Visualization*. Springer, 2014, pp. 37–59. 1

[MJG*14] MCKENZIE G., JANOWICZ K., GAO S., YANG J.-A., HU Y.: Poi pulse: A multi-granular, semantic signatures-based approach for the interactive visualization of big geosocial data. *Cartographica* (2014). 7, 13, 18

[MJR*11] MACEACHREN A. M., JAISWAL A., ROBINSON A. C., PEZANOWSKI S., SAVELYEV A., MITRA P., ZHANG X., BLANFORD J.: Senseplace2: Geotwitter analytics support for situational awareness. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2011), pp. 181–190. 6, 11, 12, 15

[MLC*13] MADDEN M., LENHART A., CORTESI S., GASSER U., DUGGAN M., SMITH A., BEATON M.: Teens, social media, and privacy. *Pew Research Center 21* (2013), 2–86. 20

[MRJ*11] MACEACHREN A. M., ROBINSON A. C., JAISWAL A., PEZANOWSKI S., SAVELYEV A., BLANFORD J., MITRA P.: Geo-twitter analytics: Applications in crisis management. *Proceedings, 25th International Cartographic Conference* (2011). 2, 6, 18

[Net24] Netlytic, [Online; accessed 2017-01-24]. `http://netlytic.org/`. 19

[NFL*14] NIU G., FAN X., LI V. O., LONG Y., XU K.: Multi-source-driven asynchronous diffusion model for video-sharing in online social networks. *IEEE Transactions on Multimedia 16*, 7 (2014), 2025–2037. 19, 20

[PHE*15] PRIETO D. F., HAGEN E., ENGEL D., BAYER D., HERNÁNDEZ J. T., GARTH C., SCHELER I.: Visual exploration of location-based social networks data in urban planning. In *Visualization Symposium (PacificVis), IEEE PacificVis* (2015), pp. 123–127. 7, 18

[PJZ*15] PANG J., JIA F., ZHANG C., ZHANG W., HUANG Q., YIN B.: Unsupervised web topic detection using a ranked clustering-like pattern across similarity cascades. *IEEE Transactions on Multimedia 17*, 6 (2015), 843–853. 19

[Plu01] PLUTCHIK R.: The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist 89*, 4 (2001), 344–350. 9

[PP11] PENNACCHIOTTI M., POPESCU A.: A machine learning approach to twitter user classification. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011* (2011). 1

[QZXS16] QIAN S., ZHANG T., XU C., SHAO J.: Multi-modal event topic model for social event analysis. *IEEE Transactions on Multimedia 18*, 2 (2016), 233–246. 19, 20

[RHD*12] ROHRDANTZ C., HAO M. C., DAYAL U., HAUG L.-E., KEIM D. A.: Feature-based visual sentiment analysis of text document streams. *ACM Transactions on Intelligent Systems and Technology (TIST) 3*, 2 (2012), 26. 9, 15

[RWD14] RIBARSKY W., WANG D. X., DOU W.: Social media analytics for competitive advantage. *Computers & Graphics 38* (2014), 328–331. 8, 18

[RZW*14] REN D., ZHANG X., WANG Z., LI J., YUAN X.: Weiboevents: A crowd sourcing weibo visual analytic system. In *Visualization Symposium (PacificVis), IEEE PacificVis Notes* (2014), pp. 330–334. 4, 5, 6, 11, 16, 20

[SDB*15] STEED C. A., DROUHARD M., BEAVER J., PYLE J., BOGEN P. L.: Matisse: A visual analytics system for exploring emotion trends in social media text streams. In *Big Data (Big Data), 2015 IEEE International Conference on* (2015), pp. 807–814. 9, 14, 15, 20

[SK13] SCHRECK T., KEIM D. A.: Visual analysis of social media data. *IEEE Computer 46*, 5 (2013), 68–75. 1

[SMER06] SHEN Z., MA K.-L., ELIASSI-RAD T.: Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1427–1439. 3

[Soc24a] Socioviz, [Online; accessed 2017-01-24]. `http://socioviz.net/`. 18, 19

[Soc24b] Socnetv, [Online; accessed 2017-01-24]. `http://socnetv.org/`. 19

[Spr24] Sprout social, [Online; accessed 2017-01-24]. `http://sproutsocial.com/`. 19

[SSMF*09] SMITH M. A., SHNEIDERMAN B., MILIC-FRAYLING N., MENDES RODRIGUES E., BARASH V., DUNNE C., CAPONE T., PERER A., GLEAVE E.: Analyzing social media networks with nodexl. In *Proceedings of the Fourth International Conference on Communities and Technologies* (2009), ACM, pp. 255–264. 3

[SWL*14] SUN G., WU Y., LIU S., PENG T.-Q., ZHU J., LIANG R.: Evoriver: Visual analysis of topic coopetition on social media. *IEEE Transactions on Visualization and Computer Graphics 20* (2014), 1753–1762. 9, 11, 12, 13, 16, 17, 20

[Syn24] Synthesio, [Online; accessed 2017-01-24]. `http://www.synthesio.com/`. 19

[Sys24] Sysomos, [Online; accessed 2017-01-24]. `http://sysomos.com/`. 19

[Tai24] Tailwind, [Online; accessed 2017-01-24]. `http://www.tailwindapp.com/`. 19

[Tal24] Talkwalker, [Online; accessed 2017-01-24]. `http://www.talkwalker.com/`. 19

[TBK*12] THOM D., BOSCH H., KOCH S., WORNER M., ERTL T.: Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Visualization Symposium (PacificVis), IEEE PacificVis* (2012), pp. 41–48. 7, 11, 13, 15, 20

[TC05] THOMAS J. J., COOK K. A. (Eds.): *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005. 12, 15

[TKE*15] THOM D., KRUGER R., ERTL T., BECHSTEDT U., PLATZ A., ZISGEN J., VOLLAND B.: Can twitter really save your life a case study of visual social media analytics for situation awareness. In *Visualization Symposium (PacificVis), IEEE PacificVis* (2015), pp. 183–190. 7, 12, 13, 15

[TSWY09] TANG J., SUN J., WANG C., YANG Z.: Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2009), KDD '09, ACM, pp. 807–816. 1

[VBW15] VEHLOW C., BECK F., WEISKOPF D.: The state of the art in visualizing group structures in graphs. pp. 021–040. http://diglib.eg.org/bitstream/handle/10.2312/eurovisstar.20151110.021-040/021-040.pdf?sequence=1. 1

[vEW14] VAN ECK N. J., WALTMAN L.: Citnetexplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics 8*, 4 (2014), 802–823. 20

[VWF09] VIEGAS F. B., WATTENBERG M., FEINBERG J.: Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics 15* (2009), 1137–1144. 8

[VWH*13] VIÉGAS F., WATTENBERG M., HEBERT J., BORGGAARD G., CICHOWLAS A., FEINBERG J., ORWANT J., WREN C.: Google+ripples: A native visualization of information flow. In *Proceedings of the 22Nd International Conference on World Wide Web* (2013), WWW '13, International World Wide Web Conferences Steering Committee, pp. 1389–1398. 2, 4, 5, 16

[WCG*16] WU Y., CAO N., GOTZ D., TAN Y.-P., KEIM D. A.: A survey on visual analytics of social media data. *IEEE Transactions on Multimedia 18* (2016), 2135–2148. 1, 20

[WDM*12] WANG X., DOU W., MA Z., VILLALOBOS J., CHEN Y., KRAFT T., RIBARSKY W.: I-si: Scalable architecture for analyzing latent topical-level information from social media data. *Computer Graphics Forum 31* (2012), 1275–1284. 8, 17

[WHMW11] WU S., HOFMAN J. M., MASON W. A., WATTS D. J.: Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011* (2011), pp. 705–714. 2

[WJS*16] WANNER F., JENTNER W., SCHRECK T., STOFFEL A., SHARALIEVA L., KEIM D. A.: Integrated visual analysis of patterns in time series and text data-workflow and application to financial data analysis. *Information Visualization 15*, 1 (2016), 75–90. 8, 18

[WLC*16] WANG X., LIU S., CHEN Y., PENG T.-Q., SU J., YANG J., GUO B.: How ideas flow across multiple social groups. In *Proc. of IEEE Visual Analytics Science and Technology* (2016). 5, 6, 10, 11, 13, 14, 20

[WLY*14] WU Y., LIU S., YAN K., LIU M., WU F.: Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1763–1772. 8, 9, 11, 17, 20

[WSJ*14] WANNER F., STOFFEL A., JÄCKLE D., KWON B. C., WEILER A., KEIM D. A., ISAACS K. E., GIMÉNEZ A., JUSUFI I., GAMBLIN T., ET AL.: State-of-the-art report of visual analysis for event detection in text data streams. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. 1

[WZSL14] WU L., ZHI Y., SUI Z., LIU Y.: Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE 9*, 5 (2014), e97010. 7, 15

[XSX*14] XIA C., SCHWARTZ R., XIE K., KREBS A., LANGDON A., TING J., NAAMAN M.: Citybeat: Real-time social media visualization of hyper-local city data. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (2014), pp. 167–170. 16

[XWW*13] XU P., WU Y., WEI E., PENG T., LIU S., ZHU J. J. H., QU H.: Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2012–2021. 8, 9, 11, 12, 20

[YJ15] YEON H., JANG Y.: Predictive visual analytics using topic composition. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction* (2015), VINCI '15, ACM, pp. 1–8. 13, 15

[YWL*14] YUAN X., WANG Z., LIU Z., GUO C., AI H., REN D.: Visualization of social media flows with interactively identified key players. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), IEEE, pp. 291–292. 6

[ZCW*14] ZHAO J., CAO N., WEN Z., SONG Y., LIN Y., COLLINS C.: Fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1773–1782. 4, 6

[ZGWZ14] ZHAO J., GOU L., WANG F., ZHOU M.: Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 203–212. 2, 9, 11, 15, 20

[Zhe15] ZHELUDEV I. N.: *When can social media lead financial markets?* PhD thesis, UCL (University College London), 2015. 18

[Zhi24] Zhiwei data, [Online; accessed 2017-01-24]. http://www.zhiweidata.com/. 19

[ZLW13] ZHANG C., LIU Y., WANG C.: Time-space varying visual analysis of micro-blog sentiment. In *Proceedings of the 6th International Symposium on Visual Information Communication and Interaction* (2013), VINCI '13, ACM, pp. 64–71. 2, 6, 7

[Zoh24] Zoho social, [Online; accessed 2017-01-24]. http://www.zoho.com/social/. 19