# Visual Analytics of Anomalous User Behaviors: A Survey

Yang Shi, Yuyin Liu, Hanghang Tong, Jingrui He, Gang Yan, and Nan Cao

**Abstract**—With the pervasive use of information technologies, the increasing availability of data provides new opportunities for understanding user behaviors. Unearthing anomalies in user behavior is of particular importance as it helps signal harmful incidents such as network intrusions, terrorist activities, and financial frauds. In this article, we survey state-of-the-art research work in visual analytics of anomalous user behaviors and classify them into four application domains, which are social interaction, travel, network communication, and financial transaction. We further examine the research work in each category in terms of data types, visualization techniques, and interactive analysis methods. We hope that our survey can provide systematic guidelines for researchers and practitioners to find effective solutions to their research problems in specific application domains. Finally, we discuss trends of academic interest over the past decades and suggest potential directions across visual analytics of these user behaviors for future research.

**Index Terms**—Anomaly detection, visual analytics, user behaviors

## 1 Introduction

The increasing accessibility of data collected from various sources provides potential opportunities for understanding user behaviors. Identifying anomalies in user behaviors is of particular interest in many application domains such as cybersecurity [1], [2], [3], urban planning [4], [5], [6], and social media [7], [8], [9], [10]. For instance, detecting rumors and tracking their spreading patterns can alert people to the risks of being influenced by misinformation [7].

However, detecting anomalous user behaviors is a challenging task as the boundary between abnormal and normal data cannot be clearly defined in most cases. Even equipped with domain knowledge, analysts may find results of automatic machine learning approaches lack contextual information to support decision-making (e.g., analysts are limited to exploring who did what when where why (5W's), and how). Visualization can address the issue by integrating human knowledge into information processing tasks. It presents anomalous patterns intuitively to decision-makers, and in the meantime involves a human-machine interaction as they explore datasets.

Our survey aims to summarize state-of-the-art research work in visual analytics of anomalous user behaviors, with the purpose of highlighting current research trends as well as future directions worth investigation. To provide systematic guidelines for researchers and practitioners finding effective solutions to their research problems, we propose to characterize the research work in this field according to their application domains. The reason is that research problems in specific application domains usually share similar solutions. Also, our work attempts to explore dominant application domains that are of interest to the visualization research community instead of those that are exhaustive or exclusive considering the great overlap between data types and visualization techniques.

In this survey, we contribute a taxonomy of visual analytics of anomalous user behaviors. To the best of our knowledge, it is the first survey that explores anomalous user behaviors from a perspective of visual analytics. Based on the data collected from specific data sources, we classify user behaviors into four categories: *social interaction* describes the communication of ideas and thoughts between people. Its data is collected from publicly accessible social platforms or private telecommunication platforms; *travel* is the physical movement of users between places containing geographic information. Its data is collected from Global Positioning System (GPS), mobile phones and base stations, etc.; *network communication* is sending and receiving information between machines via networks. Its data is collected from server logs; *financial transaction* refers to monetary flows in buying and selling, whose data is collected from system logs. Following the visual analytics pipeline of user behaviors (Fig. 1), we first extract four common data types, including text, network, spatiotemporal information, and multidimensional data. We then summarize six visualization techniques, including sequence, graph, text, geographic, chart, and glyph visualization. Finally, we propose five categories of interactive analysis methods, including tracking & monitoring, exploration & navigation, pattern discovery, knowledge externalization, and refinement & identification.

- *Y. Shi, G. Yan, and N. Cao are with the Tongji University, Shanghai, China. E-mail: {yangshi.idvx, gyan, nan.cao}@tongji.edu.cn.*
- *Y. Liu is with the Chinese University of Hong Kong, China. E-mail: yuyin.liu@outlook.com.*
- *H. Tong and J. He are with the University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA. E-mail: {htong, jingrui}@illinois.edu.*

*Manuscript received 13 May 2019; revised 27 Nov. 2019; accepted 28 Dec. 2019. Date of publication 6 Jan. 2020; date of current version 14 Mar. 2022. (Corresponding author: Yang Shi.)*
*Recommended for acceptance by Y. Tong.*
*Digital Object Identifier no. 10.1109/TBDATA.2020.2964169*

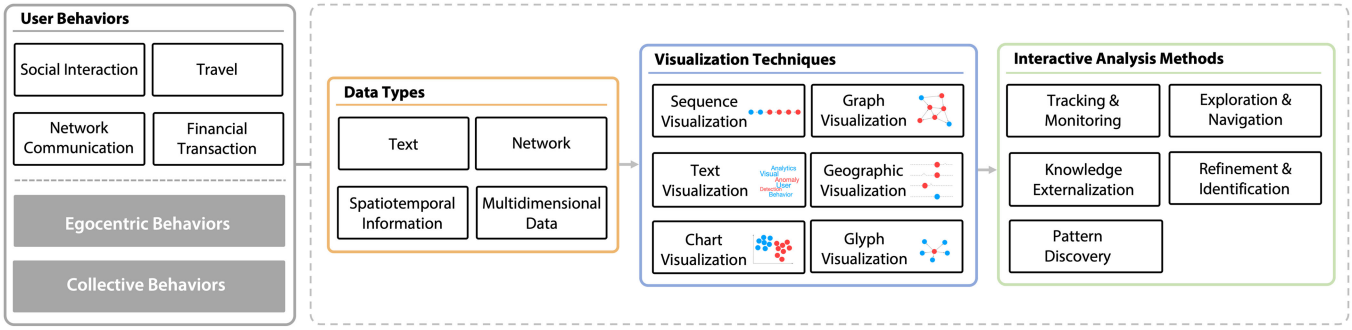


Fig. 1. Taxonomy of this survey, addressing the data type, visualization techniques, and interactive analysis methods in the visual analysis of anomalous user behaviors.

### 1.1 Scope of the Survey

To clarify the scope of the survey, we articulate the definitions of three terms used in our work: *user*, *user behaviors*, and *anomalous user behaviors*. First, a *user* is defined as "*a person who uses the functionality provided by a system*" [11]. In this case, to be classified as a user, a person uses a system in a way that the capabilities of the system fulfill the requirements or intents of the person. Second, *user behaviors* should not be confused with user-involved actions or human behaviors in this survey. Rather, user behaviors emphasize the interaction and transaction between a user and a system capable of achieving the user's requirements. Our definitions of *user* and *user behaviors* therefore rule out research works that study videos of sports events [12], patient records [13], or eye-tracking behaviors [14], as no interaction between users and machines/documents is involved. Third, *anomalous user behaviors* refers to a set of actions that break rules, deviate from expectation, occur infrequently, occur repetitively, happen for the first time or appear different from others within a specific time window. The definition of anomalies in our work shares the meaning identified in [15], "*threats and potential incidents, commonly presented as an activity that is anomalous to the standard profiles and behaviors of users and entities across time and peer group horizons*". This makes the scope of anomaly detection in our work broader than that in specific domains. For example, Chen *et al.* [16] identify data outside normal ranges of attributes as anomalies in social media, while in the field of cybersecurity, anomalies refer to malware, threats, response (e.g., malicious insiders), and targeted attacks [15].

### 1.2 Related Surveys

A few surveys that focus on analyzing user behaviors exist in recent literature. Jin *et al.* [17] categorize user behaviors in online social networks into four types including connectivity and interaction, traffic activity, mobile social behavior, and malicious behavior. Jiang *et al.* [18] classify anomalous behaviors of using web applications (e.g., Hotmail, Facebook) into four categories: traditional spam, fake reviews, social spam, and link farming. Surveys regarding visualization of user behavior data explore application domains such as urban computing [19], social media [16], [20], finance [21], and network security [22], [23]. In the field of anomaly detection, Chandola *et al.* introduce categories of anomaly detection techniques [24]. [25] and [26] examine techniques used in intrusion detection systems and for detecting graph-based anomalies, respectively. Yu *et al.* [27] investigate new categories of social media anomaly detection and review two major types of anomalies, including point and group anomalies. A recent work by Chalapathy and Chawla [28] presents a structured overview of research approaches in deep learning-based anomaly detection. Our survey covers a wider range of application domains than existing surveys. We believe that this survey will provide rich resources and useful guidelines to advance future research in the field for both researchers and practitioners.

### 1.3 Organization of the Survey

The remaining survey is organized as follows. First, we present the methodology and taxonomy used in this survey in Section 2. Sections 3, 4, 5, and 6 analyze the four user behaviors respectively using the taxonomies explained in Section 2. Analysis of each behavior follows the general visual analytics pipeline (Fig. 1). We start with identifying data types, and then visualization techniques and interactive analysis methods are discussed. This parallel structure used in Sections 3, 4, 5, and 6 provides an instruction and guideline when researchers and practitioners are given a specific type of data for analysis. We summarize each of the four sections with a short discussion. Finally, we discuss findings and trends acquired from surveying papers across different user behaviors in Section 7 and conclude our work in Section 8.

## 2 Methodology and Taxonomy

In this section, we describe our methodology of selecting papers for the topic of the survey, followed by the taxonomy of anomalous user behaviors regarding common data types, visualization techniques, and interactive analysis methods.

### 2.1 Methodology

Our interested range of publications is constrained by three conditions: user behaviors, anomaly detection, and visual analytics/visualization. We started from a core set of relevant research work known to us in advance (e.g., [4], [7], [9]), and followed references in Section 1.2 as well as papers that cite the previously identified papers. We also conducted a keyword search for papers published in visualization conferences or journals. Examples of keywords include "anomaly", "anomalous", "outlier", "abnormal", "unusual", and "rare". The research papers were read and analyzed by two researchers in visualization to affirm that they are indeed associated with the concept of anomaly described in [15]. The association

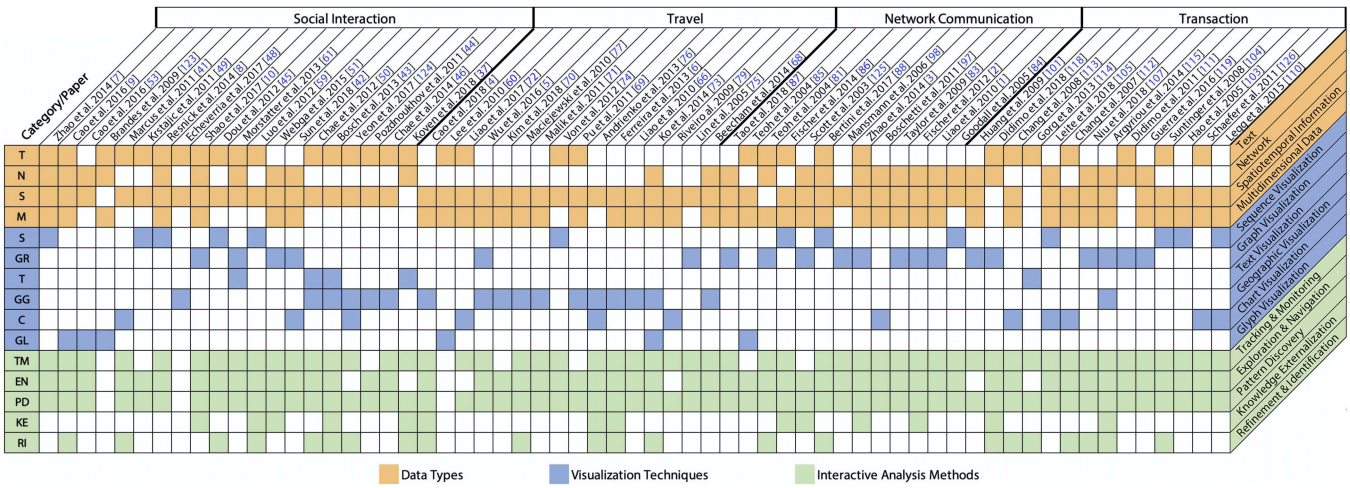


Fig. 2. Example papers regarding visualization and visual analytics of anomalous behaviors. Data Types: text (T), network (N), spatiotemporal information (S), and multidimensional data (M). Visualization Techniques: sequence (S), graph (GR), text (T), geographic (GG), chart (C), and glyph visualizations (GL). Interactive Analysis Methods: tracking & monitoring (TM), exploration & navigation (EN), knowledge externalization (KE), pattern discovery (PD), and refinement & identification (RI).

with user behaviors was expected to be seen in their application domains (e.g., Case Study in publications). Disagreements, although relatively minor, were addressed by a series of discussions between the two researchers. In the review process, the affinity diagram was used to organize the collected research work.

We also keep our exploration spectrum balanced in terms of application domains. We noticed the number of publications related to *travel* and *network communication* outnumber others. The outnumbering of *travel* probably results from the early history of visualizing spatiotemporal data (in 1869 Charles Minard produced a map to illustrate Napoleon's March to Moscow) and continuous study ever since. As for cybersecurity, the establishment of a conference for visualization of cybersecurity, IEEE Symposium on Visualization for Cyber Security (VizSec), encourages researchers to devote efforts in this field. As such, we allocated more time to searching for research work of other user behaviors comparatively. Apart from the four domains specified in this survey, we noted other user behaviors such as parallel computing performance [29], [30] and clickstream [31]. However, the number of papers found associated with these behaviors is relatively limited, especially in the field of anomaly detection. To capture possibly interesting relationships across user behaviors by maintaining a broad scope of investigation, these domains are not covered in this survey.

## 2.2 Taxonomy

Based on a literature review of more than 120 papers that relate to visual analytics of anomalous user behaviors, we summarize four categories of user behaviors, including social interaction, travel, network communication, and financial transaction. For each category, we identify common data types, visualization techniques, and interactive analysis methods. Example papers are summarized in Fig. 2.

*User Behaviors.* Anomalous user behaviors regarding social interaction, travel, network communication, and financial transaction can further be categorized into egocentric and collective behaviors. The categorization is inspired by the concepts of point and collective anomalies [24]. An *egocentric* behavior refers to the user behavior that distinguishes itself from the rest of data in anomaly detection. A *collective* behavior is a set of user behaviors that appear anomalous. When analyzing specific user behaviors categorized into collective behavior, they may appear normal on an individual basis. As egocentric and collective behaviors emphasize different aspects, specific visualization designs should be introduced. It will be discussed when analyzing visualization techniques in the following sections.

*Data Types.* A variety of data can be extracted from user behaviors across different domains. By analyzing multiple attributes of these data, we summarize four common data types including text, network, spatiotemporal information, and multidimensional data [16], [18]. A brief explanation for each data type is described as follows. *Text* provides semantic information of identities and backgrounds objects. *Network*, also called graph, consists of a set of nodes interlinked with a set of edges. A formal definition of a graph can be found in [32]. *Spatiotemporal information* captures spatial and temporal attributes of data. *Multidimensional data* uses multiple attributes to describe the properties of objects.

*Visualization Techniques.* We categorize visualization techniques that have been applied to anomalous user behaviors into sequence, graph, text, geographic, chart, and glyph visualizations.

*Sequence visualization* illustrates relations between successive events with temporal information. Anomalous sequences include spreading patterns of rumors, sudden changes in the volume of posts, and unusual business processes. Common visual representations are timeline visualization, flow visualization, and parallel coordinates.

*Graph visualization* shows structured patterns composed of nodes and edges. An anomalous graph can indicate special communication patterns in a group or communities, financial frauds conducted between employees and clients, or unauthorized network traces directed from sources to destinations. Typical graph visualizations are node-link diagrams, circular-based designs (i.e., a network topology map inside an outer ring), trees, and matrices.

*Text visualization* focuses on textual data. Anomalous text is indicated by specific keywords, topics, and sentiments extracted/abstracted from texts. Word cloud is one of the common visualization techniques for text. Text can also be combined with other visualization techniques such as flow visualization to present more contextual information.

*Geographic visualization* depicts mobility patterns of people or vehicles in geographic space. Mobility patterns include discrete as well as continuous patterns. Discrete patterns describe distribution and co-occurrence while continuous patterns depict trajectories of users when they move from one point to another. Abnormal mobility patterns are hot spots, an opposite traveling direction to most, and uncommon movement when compared to history. Heat maps and flow/bubble projections on a geographic map are used most often for visual analysis of mobility patterns.

*Chart visualization* and *Glyph visualization* represent the attributes of a multidimensional data item using a chart (e.g., x-, y-axis, color of objects) and the feature of an icon (color, size, shape), respectively. Examples of anomalies include users who only reply in a discussion board but never initiate a post and who send an unusual amount of emails at a certain time. Typical visualization techniques include 2D/3D scatter plot, bubble chart, bar chart, etc.

*Interactive Analysis Methods.* As interaction also plays an important role in visual analytics, we summarize high-level interactive analysis methods, including tracking & monitoring, exploration & navigation, knowledge externalization, pattern discovery, and refinement & identification. This typology emerges from visual analytical tasks associated with detecting anomalous user behaviors and interaction methods [33]. Analysts may mark data of interest via clicking, hovering or brushing for *tracking & monitoring*. Analysts may observe data via panning, zooming, or drill-down/roll-up functions for *exploration & navigation*. Analysts may adjust attributes of data (e.g., color, size, range) to reveal interesting patterns (*pattern discovery*). Analysts may collect, save, and extract current visualization (e.g., take a snapshot) for *knowledge externalization*. Analysts could label data with known identities (i.e., abnormal or normal data item) for *refinement & identification* of results. We believe these five categories of interactive analysis methods can provide guidance for researchers and practitioners by linking visual analytical tasks and interaction methods.

## 3 Social Interaction

Social interaction describes the communication of ideas and thoughts between people. Social interaction can be further classified into private and public interaction. *Private social interaction* behaviors include sending and/or receiving emails, making phone calls, and sending text messages between familiars on a normal basis. Examples of anomalous interaction are communication of fraudsters [34], [35] and criminals [36], [37], emailing patterns of core contributors in a working group [38], [39] and spam [40]. *Public social interaction behaviors* associated with posting/sharing/replying content on publicly accessible social platforms. Specifically, writing reviews on e-commerce platforms and editing articles in Wikipedia are also considered as public social interaction. Anomalies relate to this interaction consist of diffusion of rumors [7], [8], social bots [9], [10], and bursts of events [41], [42], [43], [44]. In particular, we include bursts of events, whose occurrence diverges from regular trends or presents a sudden block between homogeneous behaviors, as abnormal user behaviors.

We observe a few differences between private and public social interactions. The linkage between senders and receivers is not explicit in public interaction compared to one-on-one conversations in private. The information accessible on public platforms is much more than that in private settings, leading to larger volumes of data collected relevant to public behaviors. The differences can also be implied from design requirements of visual analytics tools which will be discussed in Section 3.3.

### 3.1 Data Types

Text data such as keywords, hashtags, and email contents help analysts comprehend social interaction behaviors, as it provides information including sentiment, categories, and clusters of text under a certain topic. TargetVue [9] incorporates content features to detect social bot accounts. The sudden change in the number of mentions/tags under a topic is regarded as an anomalous behavior. Echeverria *et al.* [48] discover a bot network on Twitter by solely mining the textual features of tweets. They found that the tweets of the botnet are taken directly from "Star Wars" novels. Beagle [37] allows analysts to filter contents from a filter set as well as to construct filters using keywords that are found useful during the investigation of scamming activities.

As social interaction concerns with passing, sharing, and exchanging information, networks are often seen when conversations are held between users. Follower relationships in social media, back-and-forth communication via email, and amendments made by one user in Wikipedia in response to an edit of another user are considered as network data. Gloor *et al.* [39] identify the team leader, practice leader, and practice coordinator from the visualization of social e-mail networks. These anomalous users are placed in the center of the social network and connected to multiple nodes. Fu *et al.* [38] explore small-scale email networks, where a node represents an email address, and an edge between two nodes indicates an email exchange. Analysts are able to identify specific email networks of research groups, as little communication is made across different groups. FluxFlow [7] derives anomalous user networks when exploring the process of information spreading. Indegree and outdegree are extracted based on the interaction graph of a Twitter user. These measures signal the influential power of the user.

Temporal information can be found from timestamps of microblogs, time and date of emails and calls, and days when a user appears on a forum. Locations of geo-located microblogs, locations of calls, and a terrorist network of a country are spatial data. Temporal data facilitates the analysis of communication evolution whereas spatial data shows where behavior occurs. Elzen *et al.* [34], [35] detect communication bursts using dynamic network visualization. CloudLines [49] regards sudden changes in the number of specific keywords within a period as anomalies. The keywords are collected from tweets, which arrive in data streams at non-uniform time intervals. Some visualization

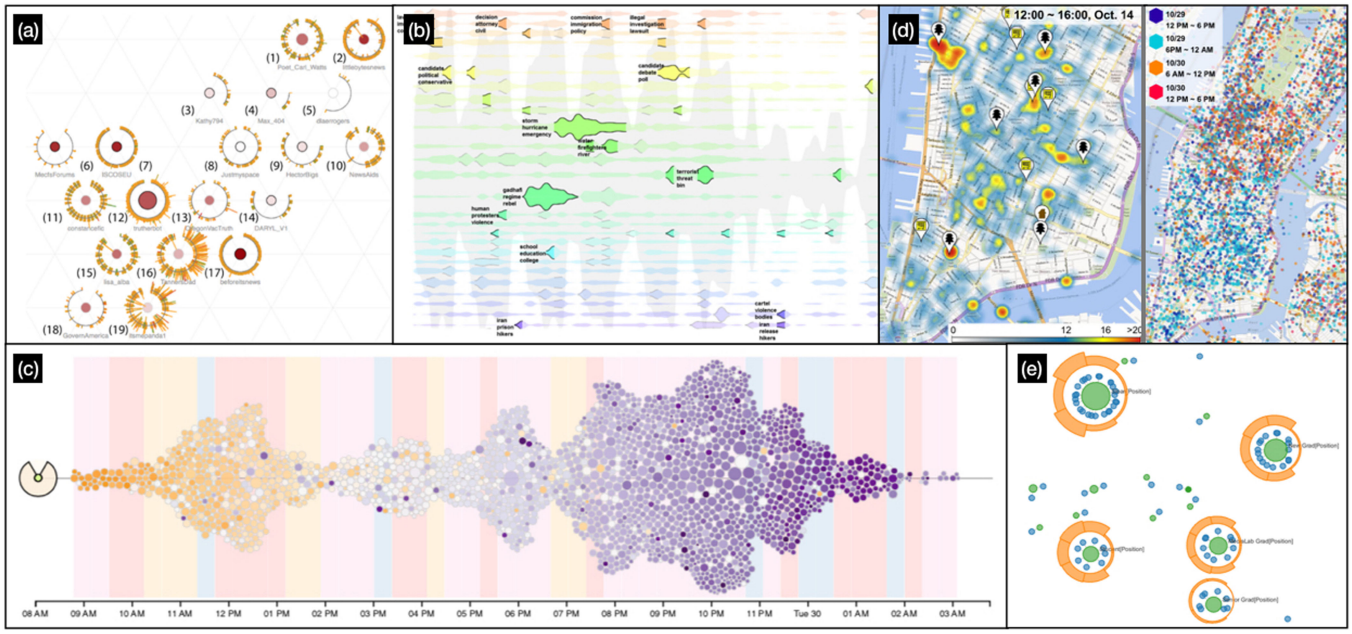


Fig. 3. Visualizations of anomalous social interaction behaviors. (a) TargetVue [9] uses a circle-based glyph visualization to encode individual users' temporal posting/reposting behaviors, anomalousness of their behaviors, and correlations between suspicious users. (b) Leadline [45] visualizes event bursts using horizontal pulse-shaped timeline visualization. (c) FluxFlow [7] shows anomalous information spreading in social media using packed circle timeline visualization. (d) Chae *et al.* [46] present public behavior response to disaster events in microblog using a heat map and hexagons on a map. (e) Mobivis [47] analyzes the calling behavior of a network consisting of university staff and students and visualizes the behavior in a node-link diagram.

works combine temporal and spatial analysis in event detection. ScatterBlogs [42], [50] detects events containing geographic information such as power outages and disasters from microblogs. In addition, messages related to the events are shown on the map.

Multidimensional data for detecting anomalous user behaviors include the length of a tweet, number of posts/emails, and average rating scores in e-commerce platforms. Webga and Lu [51] detect anomalous ratings by incorporating multidimensional data into the analysis. The multidimensional data includes the scores given by each user at the corresponding time. Rating frauds are discovered by measuring differences in average ratings and the number of rating activities in two-time windows. Cao *et al*. [9] detect anomalous users in social media by carefully selecting communication features. To investigate the interaction aspect of a social account, features such as whether users tend to communicate within a group or spread information in public, and whether users are responded from others are measured. FraudVis [52] selects ten features based on the rank of anomaly score to investigate which features contribute most to frauds on the Internet. The activity count in several time periods is one of the features that evaluate the number of views on a video website.

## 3.2 Visualization Techniques

### *3.2.1 Egocentric Behaviors*

Egocentric Social Interaction behaviors study the role of a user from his/her interaction with others. Examples of anomalous egocentric behaviors are users who only reply in a discussion board or who send an unusual amount of emails at a certain time. We observe that glyph, text, and graph visualizations are favored visual representations for egocentric behaviors.

Anomalous user behaviors can be identified via glyph visualization that is in different appearances to those of normal ones. Episogram [53] uses arrow-based and arc-based timelines to demonstrate posting and reposting activities, respectively. The two timelines can be aggregated to obtain overall tweeting behaviors. Users who always repost immediately after a message is posted are identified as arcs that always start from one end. TargetVue [9] (Fig. 3a) tackles the challenge of discovering social bots in Twitter. The circle-based glyph visualization facilitates investigation in terms of topics, sentiment, temporal dynamics of communication and its impacts, and relationships among accounts. Specifically, individual users' temporal posting/reposting behaviors, anomalousness of their behaviors, and correlations between suspicious users are encoded by behavior glyph, feature glyph, and relation glyph, respectively.

Text visualization can be used to describe egocentric communication patterns in email [54], [55]. PostHistory [54] shows the evolution of emailing patterns. It consists of two views, with one revealing the intensity of exchanged messages with each contact in a calendar view, and the other demonstrating how email addresses evolve over time in movie format. Analysts can change addresses' positions by vertical/circular/alphabetical order. Social Network Fragment [54] represents social networks in a graph where nodes are replaced by colored names of individuals. The larger the font of the name, the stronger an individual is tied to others. Viégas *et al.* [55] study changes of relationships implied from changes of keywords in email contents. The frequency and distinctiveness of keywords can be inferred from the sizes of texts, and thus anomalies such as changes of relationships (e.g., from peer to boss) can be detected.

In addition to glyph and text visualization, graph visualization, especially node-link visualization helps detect

anomalous users from their social interaction. Li *et al.* [56] explore email patterns in two graphical modes: cliques and email flows. A spambot is detected in the email flow panel when only edges originating from one node are seen. Gloor *et al.* [39], [57] investigate communication patterns of working groups in node-link visualization, and study the evolution of social structures over time in animation. Networks are drawn in a personalized mode or subject mode to identify core contributors in groups and important messages respectively. The visualization tool TeCFlow [39], [57] detects the hidden communication structure from the Enron email corpus. The hierarchical social networks uncover how Enron employees conducted collusion and fraud by emphasizing the roles of influencers, gatekeepers, and leaders. Semantic node-link views enable investigation in terms of e-mail addresses, keywords or time. Shao *et al.* [10] evaluate the extent to which an account expresses similarity to the characteristics of social bots based on diffusion patterns of tweets. In the "Hoaxy" platform, a node-link diagram is used to represent the social networks, where brighter hues indicate higher anomalous scores.

#### 3.2.2 Collective Behaviors

Collective Social Interaction behaviors are derived from users acting in a group or acting in response to each other. Anomalous collective social behaviors include temporal development of tweets, the reaction of people to special incidents, and separate group patterns of communication. Sequence, geographic and graph visualizations are used often for collective behaviors.

Sequence visualization represents the evolution of collective behaviors in various forms such as parallel coordinates and pulses/bubbles arranged along timeline visualization. Viégas *et al.* [58] visualize the revision history of Wikipedia pages in modified parallel coordinates. Each revised version of an article is represented by a vertical axis, with the axis' length indicating the length of the article. The vertical axis is divided into parts with each corresponding to revisions made by every author. By linking the axes together, a modified form of parallel coordinates shows the competition/mass deletion histories of articles. RumorLens [8] demonstrates the movement between different states of interaction associated with a rumor. The main view is a Sankey diagram. The number of people exposed to a rumor and the associated correction is illustrated with different lengths of colored segments (blue for rumors and red for corrections) in one axis. By linking different states between axes that correspond to time epochs, analysts can understand the influence of rumors and the corrections.

Pulses and bubbles arranged in a temporal sequence illustrate anomalies of collective social interaction behaviors. Major changes in the temporal development of texts are detected by highlighting unusual shapes of timelines. As one of the earliest visualizations that investigates the emergence of events, TwitInfo [41] visualizes bursts of events in a line chart. The highlighted and labeled event peaks suggest events that trigger heated discussion on Twitter. CloudLines, LeadLine, and EventRiver [45], [49], [59] detect events by relating the volume of text data extracted from online news within a period of time to the overall temporal density of keywords. Horizontal pulse-shaped timeline visualization represents event bursts, with the sizes of pulses indicating the "burstiness" of events. LeadLine (Fig. 3b) and EventRiver [45], [59] arrange vertical positions of events according to similarity of topics. FluxFlow [7] (Fig. 3c) discovers temporal trends and impacts of users in information spreading process (e.g., rumors). The main view consists of packed circles arranged along a timeline. A user's influence (i.e., the number of followers) and the corresponding anomaly score are encoded by the size and color of a circle, respectively. A user can be analyzed from three perspectives simultaneously: tweet volume, sequence, and distribution of anomalous accounts. A complimentary tree visualization demonstrates the correlation of user accounts in the diffusion process.

Geographic visualization is used to reveal events containing spatial as well as temporal references. With geographic details, anomalies can be detected from spatial intensities obtained from a collection of social interaction behaviors. Lee *et al.* [60] introduce one of the earliest works of applying spatiotemporal analysis to social media, where flows of people are represented as arrows on a map. ScatterBlogs [42], [50] employs geographic visualization for anomaly detection of topics and events as well as their spatial and temporal marks. ScatterBlogs2 [43] uses dots on a map to portray geo-located microblog posts. It differs from its previous version since there are two settings in ScatterBlogs2: a classifier creation environment and a monitoring environment. Analysts create task-tailored filters based on messages of well-understood events in the classifier creation environment, and obtain contexts of interesting events from the filter orchestration view and the time slider in the monitoring environment. Thom *et al.* [42] extract terms from messages and cluster topics as tag clouds on a zoomable map. Anomalous events are labeled and positioned on a map according to its detected location. The "Star Wars" botnet was discovered by accident when Echeverria *et al.* [48] observed sharp boundaries of the latitudinal and longitudinal position of some tweets, which were generated from bots derived from the unusual spatial distribution.

Heat map, one of the geographic visualizations, is effective at illustrating geographically-marked microblog messages. Pozdnoukhov *et al.* [44] compute heat maps from streaming tweets. The density of heat maps indicates the spatial variation of a population's response to various stimuli such as large scale sportive, political and cultural events. The differences in density between two heat maps imply the temporal evolution of events. Chae *et al.* [46] (Fig. 3d) collect a sheer volume of real-time microblog messages and mine public behavior response to disasters. A heat map and hexagons on a map identify spatiotemporal differences between crisis and normal situations.

Graph visualization including node-link and circular-based visualization uncover anomalous structures of social interaction. Perer and Shneiderman [36] emphasize the need to examine social networks systematically in *SocialAction*. The visualization tool is designed accordingly to encourage interaction with clustered node-link visualization. Analysts can quickly direct their attention to the most anomalous networks as nodes/subgroups are colored according to their ranks of anomalousness. Fu *et al.* [38]

examine small-world email networks using a combination of views. For example, stacked displays of graphs on a spherical surface visualize communication patterns between different groups. A hierarchical drawing emphasizes important nodes by placing them high in the hierarchy. MobiVis [47] (Fig. 3e) visualizes the calling behavior of a network consisting of university staff and students in a node-link diagram. The goal is to investigate information exchange and the implicit social relationship. The researchers design a "behavior ring" for user(s), which arrange events in a radial form surrounding a node. Analysts derive structural interaction from the correlation between nodes and temporal interaction from the rings.

Circular-based representation demonstrates collective social interaction behaviors in a packed visualization. Elzen *et al.* [34], [35] combine the circular hierarchical edge bundle view and massive sequence view (MSV) to detect unexpected suspicious communication patterns. The novelty of this visualization tool is that it incorporates node reordering strategies in MSV. The reordering techniques take account of closure, proximity, and similarity to ensure outliers stand out from mass data. Webga and Lu [51] project nodes (i.e., users) into a circular layout to discover rating frauds from the temporal relationship between users and items rated. The combination of the singular value decomposition diagram, re-ordered matrix representation, and temporal view reveals interesting group patterns of items. These group patterns share a similar rating history and users of similar behaviors.

### 3.3 Interactive Analysis Methods

Visual analytics of social interaction behaviors applies tracking & monitoring as one of the first steps of exploratory analysis. TwitInfo [41] tracks bursts of events in time series by highlighting event peaks in a line chart. These peaks suggest events that trigger heated discussion on Twitter. Koven *et al.* [37] multi-select summaries of email contents in the main panel to keep track of important keywords regarding scamming activities. FluxFlow [7] monitors information diffusion using multiple coordinated views. As analysts select a point in the tree view, the diffusion pattern generated by the user's reposting behavior is shown in the thread view. This interactive analysis method is usually achieved in tools with multiple coordinated views [9], [34], [47], [61], [62].

Exploration & navigation allows analysts to focus on different subranges of data flexibly. Végas *et al.* [55] design a scrolling bar, allowing analysts to review email conversations in different periods of time. TargetVue [9] enables analysts to zoom and pan in global and inspection views to locate anomalous regions. Exploration in Episogram [53] is not limited to the zooming function. Analysts can select a user of interest, and aggregate all users who perform the same posting/reposting activity. In this way, an individual's details as well as the general trend are obtained. MobiVis [47] designs a "behavior ring", from which analysts select different levels of granularity to arrange events of calling in a radial form around a node. The length of petals corresponds to the duration of selected events.

Pattern discovery is achieved in various forms of interactive analysis such as filtering. ScatterBlogs2 [43] supports generation of task-tailored filters in the classifier creation environment. In the monitoring setting, analysts can orchestrate the filters to detect anomalous users. Sorting visual objects also uncover interesting patterns. Cloudlines [49] visualizes online news events in timelines in either linear or logarithmic scale. The tool allows analysts to reconfigure visual objects via click and drag. Webga and Lu [51] detect rating frauds in the projection view, which contains two orthogonal axes inside a circle. Analysts can choose any two dimensions and a mapping method to dig out the outlier patterns. The effectiveness of pattern discovery is demonstrated with changing encoding schemes. Chae *et al.* [46] represent events detected from microblog messages with a heat map, scatters, and hexagons on a map. TargetVue [9] encodes users' action in a time sequence, anomalousness of their behaviors, and correlation to three glyph designs so that analysts acquire various perspectives of the social accounts.

Analysts may want to save the analysis results for future study. For example, documents of interest can be saved in the evidence box of the EventRiver [59] visualization tool. This function supports hypothesis evaluation and evidence exchange. Koven *et al.* [37] allow analysts to share tags created when analyzing email contents. Visualization on a website tends to have more flexible applications of knowledge externalization than stand-alone tools. After one analyzes the anomalous extents of social bots in the Hoaxy platform [10], the results can be saved into CSV files for sharing and editing.

Refinement & identification is conducted after analysts obtain a basic understanding of social interaction behaviors. LeadLine [45] associates events with the corresponding time-sensitive keywords automatically. Analysts can then annotate the events manually to provide accurate labels. There are two labeling strategies in EventRiver [59]: representative event labeling and outlier labeling. On the one hand, representative labeling is for events that contribute to the biggest cluster of a story. On the other hand, outlier labeling marks unusual events in a story. Koven *et al.* [37] emphasize tagging abilities in discoveries of anomalies. Analysts can label an account as a scammer, victim, service, or any other category. These tags can be used for creating filters and for calculation of statistics about scamming activities.

### 3.4 Discussion

Visual analytics of private social interaction behaviors related to emailing received substantial attention in the 2000s but showed a significant decrease since then. Recent research work [35], [63] is more interested in the social network structures in emailing and calling behaviors. A clear trend worth noticing is the popularity of analyzing public social interaction behaviors related to posting in social media since 2010. The volume of social media data ensures wide coverage of user behaviors including anomalous and normal behaviors. Application to the real-world is attractive from the perspective of social science, business and possibly more. We have seen many visualization tools that address event detection from massive information, information spreading, and identification of social bots. However, to the best of our knowledge, we found that only a few visualization studies [52] focus on secretive or collusive anomalous behaviors when compared to machine learning approaches [18] that

detect suspicious behaviors. Specifically, we have not seen visual analytics methods for detecting social Sybil attacks (i.e., astroturfing) [64] or private information inference [65] related to posting behaviors. We hope to see more efforts to be put in discovering anomalous behaviors conducted in a collusive, secretive manner.

## 4 Travel

Travel is the physical movement of users between places containing geographic information. Analysis of travel behaviors is meaningful for traffic monitoring, urban safety, and urban planning [4]. Travel behavior data can be collected from mobile phones and base stations, Global Positioning System (GPS), maritime search and rescue systems, and medical records. Anomalous travel behaviors differ from expected patterns indicated by individual historic records or crowd activities. Examples include irregular driving directions [4], [66], hotspots (e.g., crowded neighborhoods) [4], [5], [6], and characteristic travel patterns associated with groups of travelers [67], [68]. These anomalous behaviors can reveal potentially harmful events such as disease outbreaks and terrorist attacks.

### 4.1 Data Types

Spatiotemporal data is essential to describe when and where about users' physical motion. Spatial data consists of latitudes and longitudes, trajectories, pickup/drop-off locations, locations of base stations, etc. Temporal data includes timestamps of indoor activities, estimated time arrival, and pickup/drop-off date and time. Analysis of travel behavior usually combines both spatial and temporal data. Pu *et al.* [69] explore mobility patterns of different user groups from mobile phone data collected from each base station as well as handoff data (i.e., successive calls with different base station IDs). Spatiotemporal data related to communication includes the start time of a call, time duration, the city of the opposite side of a call, and location and direction of base stations. TelCoVis [5] explores the co-occurrence of people using telco data, which is a type of all-in-one mobile phone data containing activity records of calls, messages, and Internet usage. Data of each type of activity is comprised of timestamps, base station ID and its corresponding latitude and longitude. Kim *et al.* [70] create a visualization that helps comprehend flow patterns by analyzing the spatial distribution of non-directional discrete events over time.

Multidimensional data enriches skeletons of analysis of travel behaviors. A combination of attributes including distance traveled, speed of cars, tip amount and toll amount for taxi trips, and frequency of residents' indoor activities provides a detailed description of travelers or vehicles. Pu *et al.* [69] aggregate multidimensional data associated with base stations and mobile phone users. The data includes the total number of phone calls made by each user at each station and at all stations, in addition to spatiotemporal details. Malik *et al.* [71] evaluate the potential risks of Coast Guard search and rescue (SAR) operations for planning response actions to mitigate risks. The SAR data consists of two components: response cases and response sorties. Multidimensional data of each component contains the number of lives saved, lost, and assisted. Voila [4] extracts multidimensional features to detect abnormal incoming and outgoing taxi flows in a cell (one region is segmented into multiple cells). Examples of features include the number of vehicles that flow in and out of one cell to another.

Text associated with travel behaviors is mainly used for identification and categorization. Examples include user ID, textual messages, roam type and toll type. Pu *et al.* [69] collect the information of mobile phone ID, International Mobile Equipment Identity, city ID, roam city, roam type, and toll type to describe properties of mobile phones. These details help explain the nature of mobile phone users, i.e., travelers. Beecham *et al.* [68] categorize people into different groups in order to summarize group-cycling behaviors. Cyclists under the cycle hire scheme are classified according to age, sex, full postcode, and whether they cycle more with others or on an individual basis. Liao *et al.* [72] study resident indoor activities. These activities include not only enduring activities such as sleep, relax, watch TV, but also momentary ones such as entering home.

Network data refers to trajectories between origins and destinations. Network data is mainly used to complement spatiotemporal analysis. Ko *et al.* [73] assess flight journeys that often delay by analyzing pairs of origin and destination airports. By aggregating the number of delays for each flight journey (i.e., network), analysts detect anomalous airports and flights where prevalent delays are often found. Beecham *et al.* [68] study group-cycle journeys that link starting points and destinations.

### 4.2 Visualization Techniques

#### 4.2.1 *Egocentric Behaviors*

Egocentric Travel Behavior is individual physical movements in geographic space. An example of anomalies associated with egocentric travel behavior is an unexpected increase in time spent on indoor activities. Chart visualization is seen to represent egocentric travel behaviors.

VizTree [75] uses suffix tree visualization to indicate abnormal parts of time series by comparing with reference (i.e., normal) patterns. Anomaly detection is achieved by transforming a time series into a symbolic representation and visualizing it as a modified suffix tree. Weaver *et al.* [67] explore individual hotel visitors in a calendar view, a map view, and an arc diagram. A calendar view shows total visits on each day, with squares and circles indicating weekends and weekdays, respectively. A multi-layer map view describes paths from residences to hotels relative to railroads and rivers. By synthesizing temporal and spatial patterns observed from multiple views, analysts obtain circuitous routes taken by salesmen, cooperation between traveling merchants, and the effects of weather and seasonal variations, etc. Liao *et al.* [72] are interested in resident behaviors recorded by smart home visual systems. A heat Gantt chart view shows start time, duration, and the number of occurrences of different activities on a daily basis. By combining the heat Gantt chart with other views, activities that deviate from daily routines are detected through comparison on different daily records.

Geographic visualization is also seen for egocentric travel behaviors. A transit map displays GPS traces [66] of moving taxis in basic mode, monitoring mode, and tagging mode. Taxis are represented by glyphs on the map, with the color

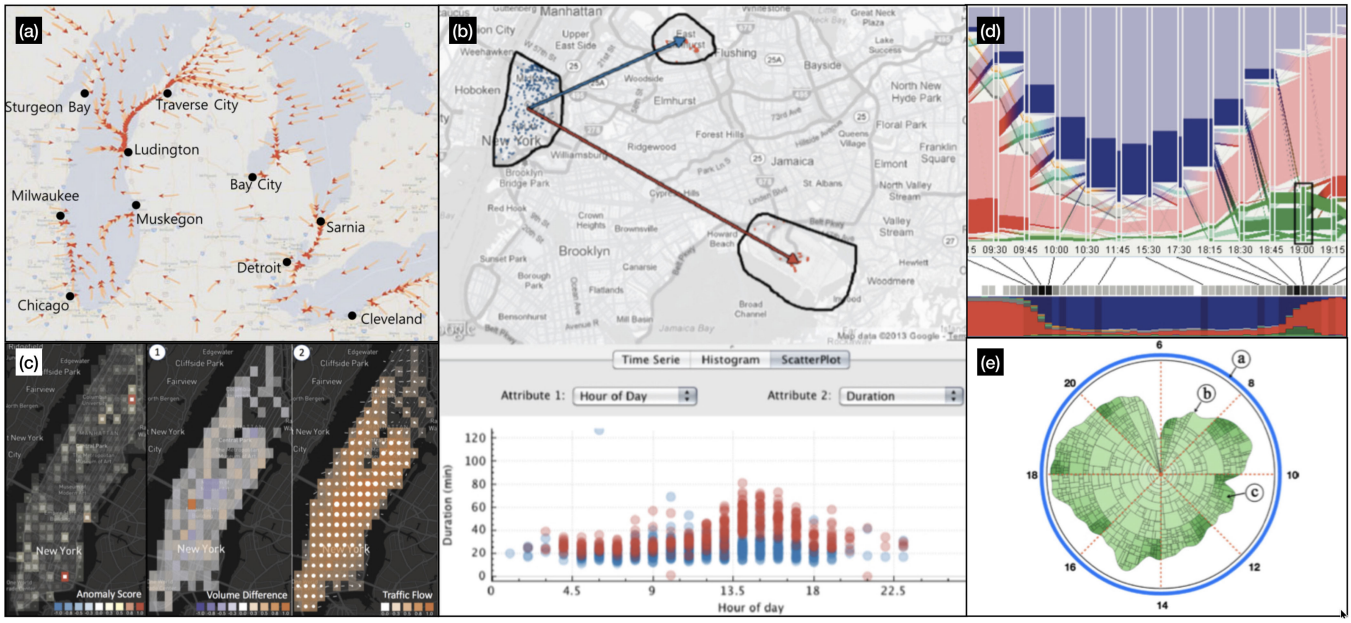


Fig. 4. Visualizations of anomalous travel behaviors. (a) Kim *et al.* [70] show origins and destinations via directions of arrows on a flow map. (b) Ferreira *et al.* [6] investigate anomalous taxi trips in New York city in multiple coordinated views consisting of a dot map and a line chart. (c) Voila [4] displays unusual traffic flows between a focal region using heap map. (d) Von Landesberger *et al.* [74] visualize different types spatiotemporal patterns by parallel coordinates. (e) Wu *et al.* [5] design a contour-based treemap to illustrate spatial and temporal characteristics of human mobility.

of glyphs dependent on whether the taxi is loaded with passengers. Egocentric anomalous travel behaviors can be a taxi in an irregular driving direction, a taxi moving at high speed, and a crowded neighborhood.

### 4.2.2 Collective Behaviors

When a collection of users move together in time and space, we say their travel behaviors are collective. Abnormal travel behaviors can be identified from regions crowded with people. As most visualization tools studying collective travel behaviors employ geographic visualization, we analyze the visual representation of travel behaviors using finer categories under geographic visualization comprised of flow maps, heat maps, and bubble/dot maps.

Flow maps represent trajectories by linking origins and destinations on a map. Andrienko and Andrienko [76] propose a framework for spatiotemporal analysis and modeling. Anomalies are found in temporal line charts displaying model residuals. Spatial flows between cells are represented by directed half-arrows whose widths are proportional to the total counts of moving objects. The flows are laid upon Voronoi maps. Trajectories of cycling patterns are shown as flows on a London city map [68]. The straight and curved end of a flow represent origin and destination respectively. Group journeys are colored red on the map whereas non-group journeys are colored blue. One of the findings is that female cyclists are more likely to make late evening journeys when cycling in groups. Kim *et al.* [70] (Fig. 4a) extract, represent, and analyze flow maps and heat maps of spatiotemporal data without the use of trajectory information. The flow map visualizes origins and destinations via directions of arrows, and the difference of flows is encoded in heat maps. Hot spots can be found with this visualization.

Heat maps display spatial densities of collective travel behaviors. Maciejewski *et al.* [77] develop an interactive visual environment to dig out hot spots in spatiotemporal data for crime analysis or surveillance syndrome. Bivariate and multivariate heat maps help detect spatiotemporal hot spots by combining height maps, colors, and contours. To analyze risks of Coast Guard search and rescue (SAR), Malik *et al.* [71] identify potential hot spots using heat maps. Risk levels of stations are indicated by the intensity of colors. The red heat map shows the time taken by stations to deploy an asset to a SAR accident while the green heat map indicates the SAR coverage. Ferreira *et al.* [6] (Fig. 4b) investigate anomalous taxi trips in New York city in multiple coordinated views of a dot map and chart visualizations. Dots on a map imply pickup and drop-off sites in a region. In the cases of Hurricane Sandy and Irene, there were virtually no dots during hurricanes, but traffic seemed to go back to normal in the following days. Voila [4] (Fig. 4c) explores taxi trips to detect sudden changes in traffic patterns. There is an anomaly detection mode giving visual cues of regional anomalies, and a context mode providing information of volume difference, traffic flow, and expected patterns in different periods of time. Unusual traffic flow between a focal region and two other places is highlighted in deep red in the heat map. Feedback from analysts can be used to update the anomaly score and thus change the color of heat map for the selected region.

We analyze other visualization techniques for travel behaviors including sequence and graph visualization. Von Landesberger *et al.* [74] (Fig. 4d) categorize spatiotemporal patterns into different types of locations according to home, work, tennis, etc. The main view is a Dynamic Categorical Data View in a varied form of parallel coordinates, which show the evolution of all types of data. Each axis of parallel coordinates indicates a point in time. When analysts select a type of data, geographic information related to the selection is plotted in the linked map, where arrows indicate physical movements of travelers. In TelCoVis [5], Wu *et al.* design a

contour-based treemap to illustrate the spatial and temporal characteristics of human mobility. By combining with heat map, matrix, and parallel coordinates, analysts are able to gain insights into the co-occurrences of human mobility and correlation of co-occurrences.

### 4.3 Interactive Analysis Methods

Analysts track and monitor data to look for anomalies. Uninteresting and expected patterns can be unmarked [75]. This improves the efficiency of detection processes and reduces false positives. TelCoVis [5] emphasizes the correlation between spatial and temporal data for exploring the co-occurrence of human mobility. When analysts hover on a sector in the contour-based treemap, all sectors corresponding to the same region will be highlighted. Moreover, analysts can mark the region of interest for exploration. Analysts can track a set of features of categoric data [74] including location, movement pattern, group membership, and group changes. The selected data instances are highlighted in the linked map view and the categoric view.

The interactive analysis associated with exploration & navigation joins separate pieces of data. Panning and altering views via scrollbars facilitate the detection of non-trivial patterns in large time series databases [75]. High-level outlooks and fine details should be accessed interchangeably when exploring travel behaviors. Different levels of aggregation in time [6], [71], [76] and space [4], [6], [68], [73] are seen in a variety of visualization tools.

Unusual travel patterns are uncovered by filtering, configuration, and encoding to various visual forms. The anomaly grading view in SHVis [72] presents anomaly scores of selected activities. Analysts click on different days and drag intervals of date to compare the activities in different periods of time. In order to analyze maritime operations and assess risks associated with the allocation of resources [71], analysts generate a combination of filters that can be applied to spatial regions and temporal plots. In addition, analysts can evaluate the effects as a result of opening/closing a station, and then determine which station is suitable for closing. Visualization can be altered in color and form to reveal anomalous patterns. Andrienko and Andrienko [76] build a framework for spatiotemporal analysis. A rich set of interactive exploration is embedded. Analysts can change the color scheme and assign colors to clusters on maps and line charts. Analysts can choose the parameters to be mapped in parallel coordinates, and adjust smoothing parameters as well as the time period for the contour-based treemap in TelCoVis [5].

Externalization of results enables analysts to keep track of important discoveries. Voila [4] includes a snapshot panel for analysts to conveniently capture the overall and detailed map views. Ferreira *et al.* [6] explore taxi trips using TaxiVis, which supports exporting query results in CSV document type, the same type of files as their input source. The visual analytics framework [76] models spatiotemporal data. Description files of models can be stored externally along with group membership of place and statistical details.

As analysts gradually develop basic knowledge, they recognize suspicious areas and integrate domain knowledge in anomaly detection. After a link is described as anomalous, the link is placed on the top of visualization while the other links become transparent [73]. In Voila [4], analysts incorporate their judgments about whether the region is anomalous. This feedback is taken into consideration in the recalculation of anomaly scores of all regions in the space.

### 4.4 Discussion

Travel receives continuous attention of researchers given that a quantity of data is available for analysis (mobile phones [74], geo-located messages [78], maritime search and rescue events [79]). Although visualization techniques used for analyzing travel behaviors are similar (i.e., geographic visualization), a rich set of interactive analysis methods is implemented in order to detect and comprehend anomalies [6], [74]. By analyzing patterns in user-specified spatial and temporal ranges, analysts study user behaviors in multiple levels of granularity, and gradually develop their understanding during interactive exploration. As more and more sensors are available in daily life, we hope to investigate a finer segmentation of travelers which would offer an accurate description of travel patterns.

## 5 Network Communication

Network communication is sending and receiving information between machines via networks. Examination of network communication has practical significance for national defense [80] and commercial enterprises [1]. Network communication behaviors include routing, network traffic, port activities, etc. Anomaly detection associated with network communication is usually concerned with cybersecurity, which is protecting computers and systems against malicious activities in a computer-related system. Anomalies are indicated by alarms as well as suspicious patterns that deviate from expectations. Investigation into these signals reveal attacks such as BGP routing instability [2], [81], virus outbreak [82], port scans [3], [83], [84], and intrusion into systems [85], [86].

### 5.1 Data Types

The identified connection between sources and destinations is seen as network data. Network data is important for detecting anomalous network communication, as it is the foundation for analyzing information exchange between machines. For example, the network connection between autonomous domains (ASes) [89] and that between subnets and hosts [90], [91] can be analyzed. Liao *et al.* [1] represent enterprise networks consisting of hosts, users, and applications as host-user-application connectivity graphs. From the graphs, the similarity of users sorted by the application can be assessed. VisAlert [80], [92] considers large-scale attack patterns between alerts and local networks. Analysts can obtain an overview of intrusion attempts and general situations by inspecting networks formed by alerts and a topology map of local network nodes.

Multidimensional data contains multiple numeric attributes to describe context information in network communication. Attack frequency, flow rates (i.e., the number of packets and bytes for a fixed period), and system load are examples of multidimensional data in the analysis of network communication behaviors. SpiralView [88] presents a connection as a list of events introduced in terms of time,

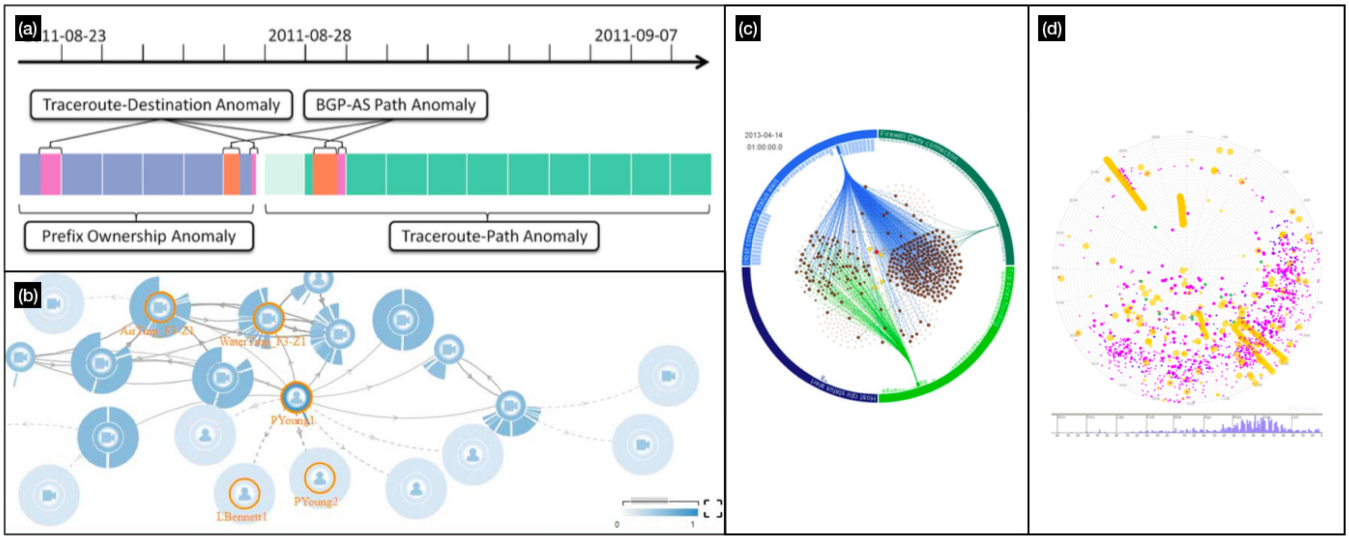


Fig. 5. Visualizations of anomalous network communication behaviors. (a) VisTracer [2] visualizes routing anomalies in traceroutes using matrix. (b) Tao *et al.* [87] design a high-order correlation graph to show collective anomalies. (c) MVSec [3] mines correlation of events attributed by what, when and where in a dandelion-metaphor using circular-based design. (d) SpiralView [88] analyzes how alarms evolve in time and detects suspicious patterns using a radar chart.

source host, application, and destination host. Details of the connection are described using multidimensional data, which are incorporated in the description of alarms. MVSec [3] uses multidimensional data including the number of connections, flow counts, and flow bytes. The statistics are combined with temporal features to explain each unit of network security data.

Spatiotemporal data of network communication associates mainly with addresses of receivers and/or senders, and temporal information of occurred activities. Spatiotemporal data provides details of timestamps and IP addresses. Investigation of spatiotemporal data is helpful for traffic monitoring, as can be seen in [93] which deals with timestamps from millisecond to year together with IP addresses from IP prefix to continents. SpiralView [88] is interested in how alarms evolve in time with the purpose of detecting periodic patterns. By inspecting alarms of the same level of attack severity, alarms can be segmented based on their temporal distribution to better understand network behaviors. VisTracer [2] visualizes destination ASes of trace-routes against time in order to explore and assess spatiotemporal patterns of occurred anomalies.

Text data type provides low-level details about connections in cyber networks. Text data can be encoded to visualization for high-level exploration, or acts as evidence for confirmation of hypothesis concerning anomalousness. Text data includes textual logs and categories of events. Erbacher *et al.* [94] represent textual log information using glyphs. Textual logs contain time, locations, and types of connection. Teoh *et al.* [85] project connections with known classes (i.e., normal, probe, DOS, U2R, and R2L) into regions in a visualization panel. Suspicious data is separate from normal data, facilitating further investigation.

## 5.2 Visualization Techniques

### 5.2.1 Egocentric Behaviors

An egocentric network communication behavior triggers alarms due to suspicious network properties of the connection between source host(s) and destination host(s). Examples of egocentric anomalous network communication behaviors are hijacking network traces by another AS, a port scan, and a high volume of traffic on a machine. Glyph and graph visualizations are used to represent egocentric behaviors.

Erbacher *et al.* [94] initiated one of the earliest visualizations to display IP addresses of alarms in a glyph-based radial form. Line glyphs surrounding a central node represent different types of connection (e.g., parallel lines indicate initial connection requests). The difference in IP addresses between the external domain and that of the monitored system is encoded in the length of line glyphs. The suspicious connection is colored red due to unexpected user activity such as timeout. Teoh *et al.* [81] inquire into Border Gateway Protocol (BGP) routing instability. Near-real-time monitoring of Internet routing is pictured as temporal line charts and glyphs, where a suspicious event detected from statistics is illustrated with a large circle in high positions and a spike in timeline.

Graph visualization, especially matrix is used to detect anomalous egocentric network communication. NVisionIP [95] detects traces of abnormal network behaviors in multiple levels of an entire class-B IP network. NVisionIP consists of a galaxy view in matrix, a small multiples view, and a machine view with bar chart. Spikes in traffic volume are seen as changes in node color in the matrix. Simple scanning attacks are discovered as clusters in the matrix, where x- and y-axis stand for subnets and hosts, respectively. VisTracer [2] (Fig. 5a) tackles large trace route data sets to distinguish legitimate routing changes and spam campaigns. The time and destination of ASes are represented by $x$- and $y$-axis in a matrix layout. Rectangular glyphs in the matrix layout are anomalies. Two nearly identical anomaly patterns in the same x-position in the matrix indicate routing anomalies in two ASes. IDSPlanet [90] uses graph visualization to support the analysis of Intrusion Detection Systems alert logs. Communication between hosts is superimposed on a circular-based design, consisting of multiple linked

components (i.e., Chrono Rings, Alert Continents, and Interactive Core) detailing port activities and raw packets.

#### 5.2.2 Collective Behaviors

Collective network communication behaviors involve more than one exchange of information between two machines or among multiple machines. Anomalous behaviors include botnet infection and periodic attacks, which are represented in graph and sequence visualizations.

Tree visualization, one of the graph visualization, helps identify anomalous network communication behaviors. Teoh *et al.* [89] examine routing behavior of BGP data. Each IP address is mapped to one pixel in a quadtree visualization to detect anomalous origin AS changes. An event is represented by a line connecting the affected IP prefix and ASes. Anomalies are revealed as an area concentrated in lines, since events that take similar paths for multiple times are suspicious. Mansmann *et al.* [93] aggregate IP addresses according to prefix, autonomous system, country and continent in treemaps based on two layout algorithms. This visualization helps monitor large-scale network data. Segments in treemaps are colored to indicate sharp changes in the number of incoming connections. Xie *et al.* [96] detect anomalous runtime behaviors by allowing analysts to explore the call stack tree representation of the executions. The anomaly detection problem is formulated as finding abnormal tree structures in a call stack forest.

Node-link diagram visualizes the structure of collective network communication. Tao *et al.* [87] (Fig. 5b) design a high-order correlation graph to show collective anomalies. When applied to software analysis, malicious attacks due to software vulnerabilities are identified as collective anomalies. In this case, a node represents each line of codes, an event represents an execution, and a correlation link represents data flow.

Circular-based visualization is also used to demonstrate collective network communication behaviors. VisAlert [80], [92] identifies critical attacks of hosts through analyzing "what, when, where" information of alerts. The alerts are allocated into segments of a ring according to the severity of attacks. "When" attribute is mapped such that the innermost ring represents the most recent activities. Inside the ring, a network topology map is used to depict the network under scrutiny. FloVis [83] observes the interaction between host pairs on either side of the monitored border. A bundle diagram displays connections between entities in a radial tree layout. Scanning activities can be detected by examining bundles directed from 9000 consecutively numbered ports to the internal host. MVSec [3] presents four coordinated views for discovering anomalies and retrieving stories behind subtle events. The event radar view (Fig. 5c) mines correlation of events attributed by what, when and where a dandelion-metaphor in a ring. Seeds (i.e., subnets) spread from the center of the dandelion stalk, which represents the only entrance to the network. Antennas (i.e., hosts) extend from the seed, giving a two-layer hierarchical structure. The seriousness of botnet infection, for instance, is indicated by the number of colored nodes in the dandelion-metaphor.

Sequence visualization uncovers abnormal trends of collective network communication. While NVisionIP [95] focuses on activities occurred on machines, its complementary tool VisFlowConnect [82] explores network flows between machines using parallel coordinates. VisFlowConnect investigates the relationship between senders and receivers. A cluster of lines originating from an external host sender indicates a virus outbreak. SpiralView [88] (Fig. 5d) analyzes how alarms evolve in time and detect suspicious patterns (e.g., alarms appear every day at the same time). The alarms are scattered dots in a radar chart, which is useful for identifying periodic intrusion patterns. The alarms are arranged starting from the center to outer parts of a circle so that more space is allocated to recent events. NStreamAware [86] analyzes a condensed heterogeneous data stream and uses a sliding slice to provide a summary for the selected period of time. The tool supports omitting and merging normal ranges so that suspicious port activities and routing behaviors, and attack patterns are revealed.

### 5.3 Interactive Analysis Methods

Detection of anomalous network communication requires tracking & monitoring. Teoh *et al.* [81] direct analysts' attention to anomalies by shading the background in gray. In the TVi [97] visual querying system, analysts select an item in the anomaly list, and then the associated temporal range is highlighted in the timeline visualization. In NVisAware [86], analysts click the star icon to store the real-time sliding slice under investigation. The events marked with star icons are added to the same view. Analysts can determine suspicious patterns from flagged and labeled events from the starred time slices. There are four coordinated views in MVSec [3]. Interaction in one view is linked to visualization in another view, which is helpful for digging hidden network attacks that are not readily recognized.

Interesting network communication behaviors are found by exploring visual elements on the same scale or in multiple levels of granularity. VisAlert [80], [92] enables panning and zooming operations of the topology map in a ring. Analysts can also configure projections onto rings by collapsing and expanding alert grouping on rings. Tao *et al.* [87] employs the direct-walk technique (i.e., a series of mouse clicks) for exploring anomalies. When an analyst notices a suspicious node, he/she clicks another node which contributes to the anomalousness of the suspicious node. That is, the analyst examines effects on the node due to other nodes. Mansmann *et al.* [93], [98] aggregate IP addresses according to prefix, autonomous system, country and continent in treemaps. Drill-down and roll-up functions can be applied for nodes of the same level of detail.

Interactive methods are used to unveil suspicious patterns of data. The filter dialogue in NVisionIP [95] restricts what data flows to be visualized. Analysts visualize network traffic according to filters based upon the combination of IP address, ports, protocols, and display type. The visual analytics tool FloVis [83] has a bundle diagram that describes network flows between a source and a destination. Analysts can loosen the bundles to find suspicious attack patterns. Additionally, analysts can choose to linearly distort points on the circle in the bundle diagram. Mansmann *et al.* [93], [98] color data in treemaps in a linear or logarithmic scale. Coloring in the logarithmic scale makes the visualization resistant to the randomness of data. Teoh *et al.* [85] use a user-directed, decision tree-based visual classification program to categorize the same

type of anomalies into one group. Analysts can interactively arrange data instances through drawing, partitioning, and appropriate coloring.

Analysts may keep a record of results for further analysis. VIAssist [99] is designed for collaborative working environments. The report builder in the visualization tool allows analysts to drag and drop graphical objects onto the current display. The annotated results can then be saved as PowerPoint or PDF files. MVSec [3] simplifies analysts' operation by offering frequently-used configuration files for anomaly detection. Analysts can export their configurations as a new configuration file.

VIAssist [99] has an expression builder and E-Diary to fulfill the refinement & identification task. Analysts can formulate a hypothesis of a suspicious activity into an expression. A catalog of expressions collects knowledge, i.e., hypotheses made by analysts during analysis. The E-Diary helps documentation of hypotheses. This encourages sharing annotations with colleagues and communication of hypotheses in a group. Analysts can annotate suspicious patterns in SpiralView [88] for long-term analysis and policy's assessment. Annotations can be an explanation for the anomalies and the action applied to the system during the analysis process.

### 5.4 Discussion

For network communication, the research interest remains relatively strong, though prior work [82], [95] that analyze this behavior are mostly published in the 2000s. Visual analytics of network communication focuses on aggregating different levels of data as well as real-time monitoring. Aggregation of data is often used to monitor high-level structures of networks and at the same time, visualizing anomalies within limited space. As data sources of audit logs and network traffic provide detailed and systematic information, attacks are often traceable to individual machines despite the fact that malicious activities originate from more than one device. Besides, the preference for real-time or near-real-time monitoring in intrusion detection [100] is emphasized, manifested by the realization of analyzing streaming data in many visualizations. These publications were motivated by the need for timely detection of malicious attacks. As computing abilities advance, we expect to see more visualization tools that handle streaming data.

## 6 Financial Transaction

Financial transaction refers to monetary flows in buying and selling. The goal is to allocate financial sources to companies or individuals. In a broad sense, stock market deals [101], credit card transactions [102], and business processes [103], [104] are under this category. Frauds are a typical type of anomalies associated with financial transactions, as people may be allured by monetary benefits to carry out illegal financial transactions. Clients may collude with employees in financial institutes in activities of money laundering, unauthorized transactions, and embezzlement [105]. Other anomalies include unexpected business processes [104], [106] and high default groups in a network of guaranteed loans [107].

### 6.1 Data Types

Spatiotemporal data describes details of location, timestamps of transactions, and time series of events. The spatiotemporal analysis is critical for detecting anomalous transactions, and thus anomaly detection often incorporates geographic locations and time series into analysis. Attributes including time of transactions [102], [104], how often a customer executes operations [108] and geographic regions [105], [109] provide a foundation for first-step analysis. For example, the Event Tunnel [104] conducts temporal correlation between seemingly isolated events, and thus business patterns and fraud patterns involving more than one individual [101] can be uncovered. Huang *et al.* [101] perform spatial correlation in addition to temporal and spectral (based on frequency) to identify suspected traders and attack plans.

Multidimensional data is often used in conjunction with spatiotemporal data to detect suspicious financial transactions. By probing into time series along with details such as the amount of money transferred [102], [105], [109], the number of financial transactions within a period of time [103], [109], and the number of the activities that are new to the user [110], analysts can gain an overview of historical financial transactions. An example of using multidimensional and spatiotemporal information is VisImpact [109]. VisImpact correlates variables of purchase quarter (i.e., temporal details), fraud amount, and fraud count to reveal relationships among important factors. Legg [110] identifies insider threats in an organization by inspecting multidimensional data including the number of times that a user performs particular tasks, the number of these activities that are new to this user and to any user in this same condition.

Network data describes relationships among entities involved in financial transactions. A network can be links between traders in trading networks [101], between entities such as people, companies and banks [111], or those between enterprises that take loan guarantees [107]. For example, high default groups are clusters of enterprises which back each other to boost their financial security, and are visualized as communities in networks [107]. Didimo *et al.* [111] analyze categorical networks that contain different types of entities to discover financial crimes. Indices such as the centrality of a node, like betweenness, and node degree are measured to indicate anomalousness.

When analyzing transaction behaviors, categories derived from text help describe the relationship between a payer and a payee [112], [113], label different types of activities conducted by employees [110], and identify the type of state changes in a business process [104]. Text data is used to distinguish between senders, intermediates, and receivers in financial transactions. Text also forms part of standard profiles that are used as references to detecting suspicious financial transactions. For example, Jigsaw [114] helps identify any linkages between people or companies relevant to financial frauds such as fictitious suppliers' invoices and systematic deletion of suppliers' invoices. These linkages are found by keyword/sentence summaries of transactions, sentiment, and word clouds of a document.

### 6.2 Visualization Techniques

#### 6.2.1 Egocentric Behaviors

An egocentric transaction is described as buying or selling behaviors conducted by an individual. An anomalous

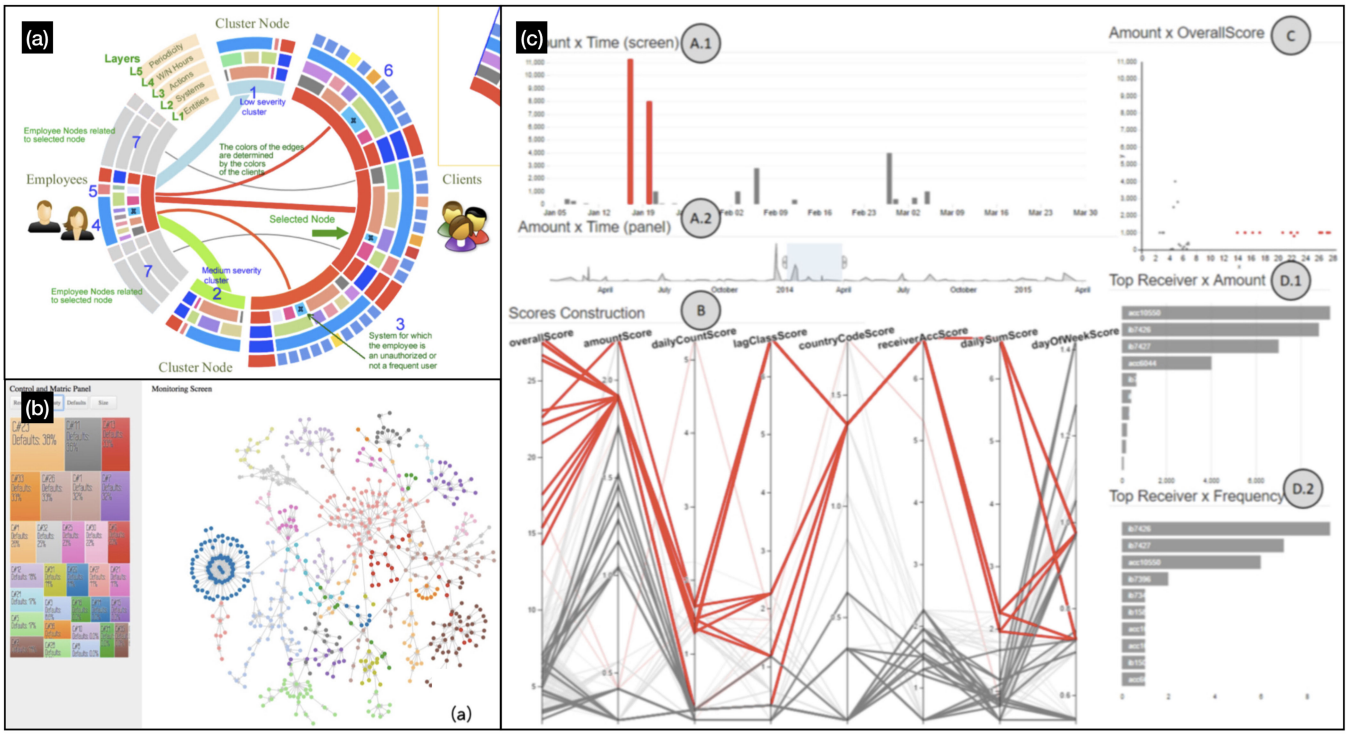


Fig. 6. Visualizations of anomalous transaction behaviors. (a) Argyriou *et al.* [115] use a multi-layer radial drawing to describe activities between employees and clients. (b) Niu *et al.* [107] assess the risk of guaranteed loans by visualizing networks of small and medium enterprises groups using a node-link visualization. (c) Leite *et al.* [105] design user-friendly views of chart visualization and parallel coordinates to help identify anomalous financial transactions.

egocentric transaction can be an unauthorized transaction or a deal worth an exceptionally high value. The detection of these behaviors mainly uses sequence visualization.

VisImpact [103], [109] organizes attributes of transactions by allocating them onto three parts/axes of a ring: left semicircle, bisector, and right semicircle. Each axis stands for an attribute of interest (e.g., region, client, fraud amount, fraud count). Suntinger *et al.* [104] display events as nodes in a cylindrical tunnel. The top view of the cylinder represents historical events, which are laid out such that more recent events are in the outer ring. Details of events are encoded by the color and size of glyphs of the Event Tunnel. Anomalous betting behaviors of a user are discovered by temporally correlating a user account in historic events to known suspicious account profiles. Argyriou *et al.* [116] study the temporal relationship of transactions between a pair of client and employee in a radar chart. The nodes in the radar chart represent transactions, which are positioned according to the time of action, pre-defined periodicity, and ordering of timelines. Events/transactions related to the same client along the radius of the radar chart are considered suspicious, as the patterns suggest the employee falsifies the client's invoices.

Graph and text visualizations are also used to demonstrate suspicious egocentric transaction behaviors. Argyriou *et al.* [115] (Fig. 6a) use a multi-layer radial drawing to describe activities between employees and clients. Each layer represents a pattern that is suspicious in different aspects (e.g., actions, systems, periodicity), with heat maps in the side view measuring anomalousness. When an employee is found to perform events that are similar to fraud patterns, a suspicious egocentric behavior is identified. Jigsaw [114] mines relationships between entities in text documents. The parallel coordinates view reveals the correlation of selected attributes (e.g., company, person). By combining with the heat map for sentiment/similarity analysis, cluster view for grouping similar documents, and document view for details, anomalous behaviors can be detected from unique text entities. Following that work, Kang *et al.* [117] study applications of Jigsaw in various situations including financial transactions. An employee's egocentric behavior of creating fictitious supplier invoices was discovered.

### 6.2.2 *Collective Behaviors*

A collective transaction behavior involves several parties in transactions and businesses. Collective transaction behaviors include a series of wire transfer and periodic transactions. Graph visualization is popular among research work for transaction behavior.

Graph visualization is popular for uncovering collective transaction anomalies. Huang *et al.* [101] develop two stages to inspect stock market security. First, market performance is evaluated using three-dimensional treemaps, with the height of blocks indicating the current price of stocks. Second, trading networks are compared against suspicious patterns in the historical database. Structured networks are regarded as collective anomalies in financial transactions. Several visual analytics tools [107], [111], [118], [119] develop categoric node-link visualizations where analysts can merge, split, define a new subgraph structure, cluster nodes according to a top-down or bottom-up paradigm, and adjust the size of nodes using a chosen measurement. Users edit networks interactively to discover communities,

signaling suspicious financial transactions. Didimo *et al.* [111] detect financial activity networks such as money laundering by illustrating entities involved in financial transactions with nodes. The entities include banks, companies, persons, bank accounts, transactions, and reports filing. Edges between nodes represent semantic connections. For instance, two disjoint clusters that indicate fraudulent patterns are revealed after clustering. The level of depth of a cluster reflects the extent of criticism of illegal activity. Niu *et al.* [107] (Fig. 6b) assess the risk of guaranteed loans by visualizing networks of small and medium enterprises which back each other and thus more exposed to default. Anomalies, i.e., high default groups, are identified as communities in the network using a node-link visualization. A complementary treemap supports the navigation of labels/categories and presentation of default rates.

Chart and sequence visualizations are also used to detect collective financial transaction behaviors. WireVis [112], [113] uses multiple coordinated chart visualization to analyze suspicious wire transfers between a payer to a payee via a chain of intermediaries. The overall trends of activities and individual transactions are represented by strings and beads respectively in an x-y plot of transaction value against time. Suspicious financial transactions are the ones relevant to a keyword that is only found in the second half of the year, and those that are of much higher value than others. Leite *et al.* [105] (Fig. 6c) design user-friendly views of chart visualization and parallel coordinates to help identify the anomalous connection between the amount and the suspicious transactions. If anomaly scores of transactions deviate from normal ranges, the days that contain at least one suspicious transaction are highlighted in red.

### 6.3 Interactive Analysis Methods

Analysts track suspicious data by highlighting and correlating relevant data. The visual analytics tool EVA [105] computes the overall anomaly scores and sub-scores according to various standards. If the overall score of transactions exceeds a threshold, the transactions are highlighted in red in the parallel coordinates view. Also, selection in another coordinated chart highlights associated transactions and grays out others in the parallel coordinates view. When analysts click a node of interest, relevant data that are originally not visualized is displayed [111]. This helps analysts not only discover interesting features that are not apparent in one view, but also identify different relationships between data instances. A similar operation is seen in [115], where the selection of one node adds related employees (i.e., nodes) into the visualization. Thus, frauds carried out by two or more employees can be tracked.

VisImpact [103], [109] supports simultaneous browsing and navigation of multiple nodes. Details of a single node representing a transaction record can be obtained using the drill-down function. To analyze transactions of an account, they can be aggregated in days, weeks, or months in WireVis [112], [113]. Zooming is enabled in the heat map and temporal chart view. One can also drill down to individuals and compare their records against each other in WireVis. Network Explorer [119] includes an overview and an egocentric mode that detects important clusters and individual nodes, respectively. In the overview mode, analysts can navigate to one cluster and compute sub-communities on demand. In the egocentric mode, analysts navigate nodes using the direct-walk from a starting point.

Pattern discovery is often used to help identify anomalous behaviors. Filtering in WireVis [112], [113] is conducted using a set of keywords and criteria like the amounts of words. Analysts select reasonably sized subsets for re-clustering to generate clusters that exhibit interesting features. Furthermore, the color scheme is chosen depending on the characteristic (e.g., sequential or diverging) of the measurement in the heat map. Jigsaw [114] allows involvement in defining clusters of text documents, removing false positives, adjusting the number of words shown, and reordering the entity list. Dragging, merging, and splitting visual elements are often seen in node-link visualization [107], [111], [118], [119]. To discover the tax evasion behaviors [118], analysts can merge and split node-link representation. A selection of subgraphs is ranked according to criteria such as the total amount of economic transactions and the risk index. Also, analysts can define and draw suspicious graph patterns using pre-defined operators.

A few visualization tools support exporting analyzed results. For example, the Event Tunnel [104] contains a snapshot management console that captures the current state and configuration. Argyriou *et al.* [115], [116] design the exporting function in the visual analytics tool for detecting occupational frauds. The results of ranking in anomalousness can be exported in separate log files. The visualization containing suspicious transaction patterns can be stored for post-analysis.

Visual analytics involves domain knowledge into the process of anomaly detection. Analysts are enabled to reassign labels of the "structure hole spanner" during interactive exploration [107]. The structure hole spanner interlinks different communities in a network, which can be modified through merging and splitting operations. High default groups are found to be associated with these labels. In TAXNET [118], analysts can define graph patterns based on their understanding of tax evasion frauds. Textual labels are attached to graphs to provide a description of rules for nodes (i.e., taxpayers) or edges (i.e., relationship).

### 6.4 Discussion

Visualization works regarding anomalous financial transaction behaviors modernize visual methods in the field of finance. For example, EVA [105] integrates human decisions into the analysis of frauds into the existing alert system. In recent years, we have seen an increasing number of visualization tools designed for detecting suspicious users involved in financial transactions. However, by comparing the average number of citations between user behaviors, the overall research interest in financial transactions is less than those in travel behaviors, for example. Privacy issues can largely limit the resources available for research. Having said that, we are hoping to see more in-depth collaboration between academic researchers and financial institutes to resolve transaction frauds by exploring and understanding fraudsters' behaviors.

## 7 Discussion and Outlook

In this section, we discuss our findings regarding data types, visualization techniques, and interactive analysis methods

across the four user behaviors. We also provide takeaways to facilitate future visual analytics of user behaviors.

### 7.1 Data Types

Application of *multidimensional data* to anomaly detection can be found across the four behaviors. It offers a variety of features for detecting anomalous behaviors and is often used in conjunction with other data types. *Text* is an important data type for detecting abnormal social interaction behaviors, whereas text is a complement in the analysis of other user behaviors. Text provides information about the identities and backgrounds of objects involved, which is used to categorize objects. *Network* is used frequently in the analysis of network communication as well as social interaction behaviors. Links exist in cyber networks between sources and destinations, and social networks between senders and receivers. *Spatiotemporal information* enriches skeletons of analysis by incorporating contextual information of users' travel behaviors. Besides, temporal analysis is often incorporated in detecting anomalous transactions and social interaction behaviors.

Analysis based on data types helps indicate overlapping areas between user behaviors, which is a signal of borrowing analytics approaches from other behaviors. For example, the exploration of rating behaviors in online e-commerce stores is similar to that of network security problems. Sensitivity to time-critical behaviors in anomaly detection is emphasized in [51], in which streaming data is processed. The network between sources and destinations is found in network communication, whilst network between users and items is also important for discovering rating frauds.

*Takeaways*. Incorporating multiple data types for visual analytics is gaining importance among recent researcher work. Since anomaly detection problems often encounter unknown ill-defined anomalies and heterogeneous data, usage of multiple data types can create a relatively thorough picture for investigation.

### 7.2 Visualization Techniques

Among *graph visualization*, node-link diagram is mostly used in social interaction, financial transaction, and network communication. Node-link diagram is advantageous in its traceability from one node to the other. It is capable of tracing down to abnormal individuals from email and call records, to individual machines in malicious cyber-attacks, and to a pair of employee and client in financial frauds. *Text visualization* is favored in the analysis of public social interaction behaviors such as posting. These visualization tools are usually equipped with views containing texts, which support interactive exploration and affirmation of suspicious events or users. For example, to complement inspection of microblogs, original messages and keywords are often found in a table format or tag clouds [9], [120].

Detection of abnormal transaction behaviors uses *sequence visualization* such as parallel coordinates. Variations of relationships between subsequent events can be tracked down through changes of linkage between two successive axes, which suggest suspicious transactions may occur. Varied configurations of parallel coordinates include radar chart and Sankey diagram. To illustrate social interaction behaviors, changes of heights and size of bubbles in timeline visualization are used to encode sudden and/or important changes in the volume of keywords. *Geographic visualization* is often used to represent travel behaviors as it has the advantage of illustrating two-dimensional physical movement. Flows and bubbles projected on a map show differences in traveling directions and spatial densities of distribution. Heat map is popular to demonstrate spatial densities of humans and vehicles, as it minimizes visual occlusion that may happen in flows/bubbles projection on maps. *Chart visualization* is effective in illustrating well-understood anomalies as long as dimensions of the displays are selected properly.

When comparing visualization techniques that are applied to egocentric and collective behaviors, we found glyph visualization is suitable for visualizing egocentric behaviors as differences in individuals' roles can be identified more effectively. Comparatively visualization of collective behaviors takes a variety of representations. To better explain, we use an example in social media where the same user behavior results in problems viewed from egocentric and collective perspectives, respectively. Both Episogram [53] and FluxFlow [7] analyze retweeting behaviors on Twitter. Episogram considers whether a Twitter account is anomalous by comparing one's individual retweeting patterns with others'. A user is represented as a glyph, which is later found to be used as a typical visualization for egocentric behavior form. FluxFlow, on the other hand, emphasizes the information diffusion process and visualizes the temporal evolution of a group of retweeted microblogs using packed colored circles.

*Takeaways*. Node-link diagram has long been a popular choice of visualizing anomalous user behaviors. It is still one of the favored techniques as it is powerful to demonstrate an overall structure as well as detailed information when incorporated with rich interactive analysis methods. Circular-based designs are gaining attention among researchers due to its ability to show connections in a packed visualization, where the hierarchical structure is displayed using bundles and tree layout inside the ring. Heat map is increasingly used when compared to flows/bubbles/3D projection on a map. The reason is that heat map visualizes large-scale data containing geographic information, and in the meantime encodes variables such as anomaly degrees without occlusion.

### 7.3 Interactive Analysis Methods

*Exploration & navigation* has been the most popular interactive analysis method in visual analytics of anomalous user behaviors. Most visualization tools enable users to gain a high-level summary of large-scale data first and then to drill down to details upon request. The fact that exploration & navigation is used most is consistent with the characteristic of data visualization, i.e., quick interpretation of a sizable amount of data. It also agrees with Shneiderman's Visual Information Seeking Mantra: "Overview first, zoom and filter, then details-on-demand" [121]. The second most popular interactive analysis task is *tracking & monitoring*. As the papers surveyed are related to anomaly detection, keeping track of suspicious spots is important during interactive

exploration. Analysts highlight data of interest to present its correlation in the coordinated views, which helps form a picture of where anomalies originate from. *Pattern discovery* is also frequently used. It allows analysts to focus on data within a certain period of time, related to specific texts, or within specific ranges. During the interactive process, the visual representation of data changes accordingly. These updates of one's understanding drive analysts to construct accurate hypotheses of anomalies.

We observe trends in utilizing interactive analysis methods in different user behaviors. Visualization works that study travel behaviors often incorporate *exploration & navigation* in map visualization. The reason is that panning on a map is seen often when tracking physical movement [4], [76]. Pattern discovery illustrates more than one abnormal feature of anomalies by changing the color spectrum and representing traveling patterns in various forms on a map [6], [46]. Also, filtering by keywords is seen in social interaction [9], [36], [51], [122] where textual contents are important for determining anomalies. *Knowledge externalization* is usually seen in network communication [3], [99] and transactions [112], [115]. This interactive analysis method enables the processed results to be outputted for further analysis and validation with domain experts. *Refinement & identification* is increasingly used by researchers. This method goes beyond the definition of interaction methods [33] because adjustments in the parameters of anomaly detection algorithms are allowed (e.g., in Filter technique), resulting in the modification of visual representation. Several publications allow analysts to adjust parameters in constructing queries [6], [43], changing thresholds of anomalies [51], [74], and updating feedback in anomalies [4].

*Takeaways*. As more and more research work is leveraging interactive machine learning for anomaly detection, refinement & identification is of interest to the visualization community. Refinement & identification can support visual analytics by involving human perception and interpretation into the computation process of anomaly detection, which is a deeper level of computer-human interaction.

## 8 Conclusion

With the increasing accessibility of data collected from various sources, many researchers have realized the importance of applying visual analytics to understand anomalous user behaviors. To facilitate the investigation pursuit, we present a survey of visual analytics of anomalous user behaviors. We analyze relevant state-of-the-art papers according to the proposed taxonomies. Our survey suggests trends and preferences in data types, anomaly detection techniques, visualization techniques, and interactive analysis methods. With these findings, we also highlight potential research directions. We believe that our work can shed light on understanding and analyzing anomalous user behaviors using visual analytics approaches.

## Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities in China and NSFC (61802283 and 61602306). Hanghang Tong is partially supported by NSF (1939725 and 1715385). Jingrui He is partially supported by NSF (1947203 and 1813464).

## References

[1] Q. Liao, A. Striegel, and N. Chawla, "Visualizing graph dynamics and similarity for enterprise network security and management," in *Proc. 7th Int. Symp. Vis. Cyber Security*, 2010, pp. 34–45.

[2] F. Fischer, J. Fuchs, P.-A. Vervier, F. Mansmann, and O. Thonnard, "Vistracer: A visual analytics tool to investigate routing anomalies in traceroutes," in *Proc. 9th Int. Symp. Vis. Cyber Security*, 2012, pp. 80–87.

[3] Y. Zhao, X. Liang, X. Fan, Y. Wang, M. Yang, and F. Zhou, "MVSec: Multi-perspective and deductive visual analytics on heterogeneous network security data," *J. Vis.*, vol. 17, no. 3, pp. 181–196, 2014.

[4] N. Cao, C. Lin, Q. Zhu, Y.-R. Lin, X. Teng, and X. Wen, "Voila: Visual anomaly detection and monitoring with streaming spatio-temporal data," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 23–33, Jan. 2018.

[5] W. Wu *et al.*, "Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 935–944, Jan. 2016.

[6] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2149–2158, Dec. 2013.

[7] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins, "# fluxflow: Visual analysis of anomalous information spreading on social media," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1773–1782, Dec. 2014.

[8] P. Resnick, S. Carton, S. Park, Y. Shen, and N. Zeffer, "Rumorlens: A system for analyzing the impact of rumors and corrections in social media," in *Proc. Comput. Journalism Conf.*, 2014, pp. 10 121–0701.

[9] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin, "Targetvue: Visual analysis of anomalous user behaviors in online communication systems," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 280–289, Jan. 2016.

[10] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots,", *Nat. Commun.*, vol. 9, no. 1, pp. 4787–4795, 2018.

[11] M. Glinz, "A glossary of requirements engineering terminology," *Standard Glossary Certified Professional Requirements Eng. Studies Exam Version*, vol. 1, pp. 9–24, 2011.

[12] M. Stein *et al.*, "Bring it to the pitch: Combining video and movement data to enhance team sport analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 13–22, Jan. 2018.

[13] N. Cao, D. Gotz, J. Sun, and H. Qu, "Dicon: Interactive visual analysis of multidimensional clusters," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2581–2590, Dec. 2011.

[14] G. Andrienko, N. Andrienko, M. Burch, and D. Weiskopf, "Visual analytics methodology for eye movement studies," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2889–2898, Dec. 2012.

[15] G. Sadowski, A. Litan, T. Bussa, and T. Phillips, "Market guide for user and entity behavior analytics," 2018. [Online]. Available: https://www.cbronline.com/wp-content/uploads/dlm_uploads/2018/07/gartner-market-guide-for-ueba-2018-analyst-report.pdf

[16] S. Chen, L. Lin, and X. Yuan, "Social media visual analytics," in *Computer Graphics Forum*, vol. 36, Hoboken, NJ, USA: Wiley, 2017, pp. 563–587.

[17] L. Jin, Y. Chen, T. Wang, P. Hui, and A. V. Vasilakos, "Understanding user behavior in online social networks: A survey," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 144–150, Sep. 2013.

[18] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious behavior detection: Current trends and future directions," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 31–39, Jan./Feb. 2016.

[19] Y. Zheng, W. Wu, Y. Chen, H. Qu, and L. M. Ni, "Visual analytics in urban computing: An overview," *IEEE Trans. Big Data*, vol. 2, no. 3, pp. 276–296, Sep. 2016.

[20] Y. Wu, N. Cao, D. Gotz, Y.-P. Tan, and D. A. Keim, "A survey on visual analytics of social media data," *IEEE Trans. Multimedia*, vol. 18, no. 11, pp. 2135–2148, Nov. 2016.

[21] S. Ko *et al.*, "A survey on visual analysis approaches for financial data," in *Computer Graphics Forum*, vol. 35, Hoboken, NJ, USA: Wiley, 2016, pp. 599–617.

[22] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 8, pp. 1313–1329, Aug. 2012.

[23] V. Lavigne and D. Gouin, "Visual analytics for cyber security and intelligence," *J. Defense Model. Simul.*, vol. 11, no. 2, pp. 175–199, 2014.

[24] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surveys*, vol. 41, pp. 15:1–15:58, 2009.

[25] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, pp. 3448–3470, 2007.

[26] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Mining Knowl. Discovery*, vol. 29, pp. 626–688, 2014.

[27] R. Yu, H. Qiu, Z. Wen, C. Lin, and Y. Liu, "A survey on social media anomaly detection," *ACM SIGKDD Explorations Newsletter*, vol. 18, no. 1, pp. 1–14, 2016.

[28] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019. [Online]. Available: https://arxiv.org/pdf/1901.03407

[29] C. Muelder, B. Zhu, W. Chen, H. Zhang, and K.-L. Ma, "Visual analysis of cloud computing performance using behavioral lines," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 6, pp. 1694–1704, Jun. 2016.

[30] D. K. Osmari, H. T. Vo, C. T. Silva, J. L. Comba, and L. Lins, "Visualization and analysis of parallel dataflow execution with smart traces," in *Proc. 27th SIBGRAPI Conf. Graphics Patterns Images*, 2014, pp. 165–172.

[31] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson, "Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 321–330, Jan. 2017.

[32] R. Balakrishnan and K. Ranganathan, *A Textbook of Graph Theory*. Berlin, Germany: Springer, 2012.

[33] J. S. Yi, Y. ah Kang, and J. T. Stasko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1224–1231, Dec. 2007.

[34] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, "Reordering massive sequence views: Enabling temporal and structural analysis of dynamic networks," in *Proc. IEEE Pacific Vis. Symp.*, 2013, pp. 33–40.

[35] S. van den Elzen, D. Holten, J. Blaas, and J. J. van Wijk, "Dynamic network visualization withextended massive sequence views," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 8, pp. 1087–1099, Aug. 2014.

[36] A. Perer and B. Shneiderman, "Balancing systematic and flexible exploration of social networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 5, pp. 693–700, Sep./Oct. 2006.

[37] J. Koven, C. Felix, H. Siadati, M. Jakobsson, and E. Bertini, "Lessons learned developing a visual analytics solution for investigative analysis of scamming activities," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 225–234, Jan. 2019.

[38] X. Fu, S.-H. Hong, N. S. Nikolov, X. Shen, Y. Wu, and K. Xuk, "Visualization and analysis of email networks," in *Proc. 6th Int. Asia-Pacific Symp. Vis.*, 2007, pp. 1–8.

[39] P. A. Gloor and Y. Zhao, "Tecflow-a temporal communication flow visualizer for social networks analysis," in *Proc. ACM CSCW Workshop Social Netw.*, 2004, vol. 6.

[40] C. Muelder and K.-L. Ma, "Visualization of sanitized email logs for spam analysis," in *Proc. 6th Int. Asia-Pacific Symp. Vis.*, 2007, pp. 9–16.

[41] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: Aggregating and visualizing microblogs for event exploration," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 227–236.

[42] D. Thom, H. Bosch, S. Koch, M. Wörner, and T. Ertl, "Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages," in *Proc. IEEE Pacific Vis. Symp.*, 2012, pp. 41–48.

[43] H. Bosch *et al.*, "Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2022–2031, Dec. 2013.

[44] A. Pozdnoukhov and C. Kaiser, "Space-time dynamics of topics in streaming text," in *Proc. 3rd ACM SIGSPATIAL Int. Workshop Location-Based Social Netw.*, 2011, pp. 1–8.

[45] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou, "Leadline: Interactive visual analysis of text data through event identification and exploration," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, 2012, pp. 93–102.

[46] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, and D. S. Ebert, "Public behavior response analysis in disaster events utilizing visual analytics of microblog data," *Comput. Graphics*, vol. 38, pp. 51–60, 2014.

[47] Z. Shen and K.-L. Ma, "Mobivis: A visualization system for exploring mobile data," in *Proc. IEEE Pacific Vis. Symp.*, 2008, pp. 175–182.

[48] J. Echeverria and S. Zhou, "Discovery, retrieval, and analysis of the'star wars' botnet in twitter," in *Proc. IEEE/ACM Int. Conf. Advances Social Netw. Anal. Mining*, 2017, pp. 1–8.

[49] M. Krstajic, E. Bertini, and D. Keim, "Cloudlines: Compact display of event episodes in multiple time-series," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 12, pp. 2432–2439, Dec. 2011.

[50] J. Chae *et al.*, "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 143–152.

[51] K. Webga and A. Lu, "Discovery of rating fraud with real-time streaming visual analytics," in *Proc. IEEE Symp. Vis. Cyber Security*, 2015, pp. 1–8.

[52] J. Sun *et al.*, "Fraudvis: Understanding unsupervised fraud detection algorithms," in *Proc. IEEE Pacific Vis. Symp.*, 2018, pp. 170–174.

[53] N. Cao, Y.-R. Lin, F. Du, and D. Wang, "Episogram: Visual summarization of egocentric social interactions," *IEEE Comput. Graphics Appl.*, vol. 36, no. 5, pp. 72–81, Sep./Oct. 2016.

[54] F. B. Viégas, D. Boyd, D. H. Nguyen, J. Potter, and J. Donath, "Digital artifacts for remembering and storytelling: Posthistory and social network fragments," in *Proc. 37th Annu. Hawaii Int. Conf.*, 2004, p. 10.

[55] F. B. Viégas, S. Golder, and J. Donath, "Visualizing email content: portraying relationships from conversational histories," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2006, pp. 979–988.

[56] W.-J. Li, S. Hershkop, and S. J. Stolfo, "Email archive analysis through graphical visualization," in *Proc. ACM Workshop Vis. Data Mining Comput. Security*, 2004, pp. 128–132.

[57] P. A. Gloor, S. Niepel, and Y. Li, "Identifying potential suspects by temporal link analysis," in *MIT CCS Working Paper*, 2006.

[58] F. B. Viégas, M. Wattenberg, and K. Dave, "Studying cooperation and conflict between authors with history flow visualizations," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2004, pp. 575–582.

[59] D. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. Keim, "Eventriver: Visually exploring text collections with temporal references," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 1, pp. 93–105, Jan. 2012.

[60] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for twitter-based GEO-social event detection," in *Proc. 2nd ACM SIGSPATIAL Int. Workshop Location Based Social Netw.*, 2010, pp. 1–10.

[61] F. Morstatter, S. Kumar, H. Liu, and R. Maciejewski, "Understanding twitter data with tweetxplorer," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 1482–1485.

[62] F. B. Viégas and M. Smith, "Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces," in *Proc. 37th Annu. Hawaii Int. Conf.*, 2004, p. 10.

[63] D. Redondo, A. Sallaberry, D. Ienco, F. Zaidi, and P. Poncelet, "Layer-centered approach for multigraphs visualization," in *Proc. IEEE 19th Int. Conf. Inf. Vis.*, 2015, pp. 50–55.

[64] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *Proc. IEEE Symp. Security Privacy*, 2008, pp. 3–17.

[65] R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Preventing private information inference attacks on social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1849–1862, Aug. 2013.

[66] Z. Liao, Y. Yu, and B. Chen, "Anomaly detection in GPS data based on visual analytics," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, 2010, pp. 51–58.

[67] C. Weaver, D. Fyfe, A. Robinson, D. Holdsworth, D. Peuquet, and A. M. MacEachren, "Visual exploration and analysis of historic hotel visits," *Inf. Vis.*, vol. 6, no. 1, pp. 89–103, 2007.

[68] R. Beecham and J. Wood, "Characterising group-cycling journeys using interactive graphics," *Transp. Res. Part C: Emerg. Technol.*, vol. 47, pp. 194–206, 2014.

[69] J. Pu, P. Xu, H. Qu, W. Cui, S. Liu, and L. Ni, "Visual analysis of people's mobility pattern from mobile phone data," in *Proc. Vis. Inf. Commun.-Int. Symp.*, 2011, Art. no. 13.

[70] S. Kim, S. Jeong, I. Woo, Y. Jang, R. Maciejewski, and D. S. Ebert, "Data flow analysis and visualization for spatiotemporal statistical data without trajectory information," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 3, pp. 1287–1300, Mar. 2018.

[71] A. Malik, R. Maciejewski, B. Maule, and D. S. Ebert, "A visual analytics process for maritime resource allocation and risk assessment," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, 2011, pp. 221–230.

[72] Z. Liao *et al.*, "A visual analytics approach for detecting and understanding anomalous resident behaviors in smart healthcare," *Appl. Sci.*, vol. 7, no. 3, 2017, Art. no. 254.

[73] S. Ko *et al.*, "Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, 2014, pp. 83–92.

[74] T. Von Landesberger, S. Bremm, N. Andrienko, G. Andrienko, and M. Tekusova, "Visual analytics methods for categoric spatio-temporal data," in *Proc. IEEE Conf. Vis. Analytics Sci. Technol.*, 2012, pp. 183–192.

[75] J. Lin, E. Keogh, and S. Lonardi, "Visualizing and discovering non-trivial patterns in large time series databases," *Inf. Vis.*, vol. 4, no. 2, pp. 61–82, 2005.

[76] N. Andrienko and G. Andrienko, "A visual analytics framework for spatio-temporal analysis and modelling," *Data Mining Knowl. Discovery*, vol. 27, no. 1, pp. 55–83, 2013.

[77] R. Maciejewski *et al.*, "A visual analytics approach to understanding spatiotemporal hotspots," *IEEE Trans. Vis. Comput. Graphics*, vol. 16, no. 2, pp. 205–220, Mar./Apr. 2010.

[78] J. Thomas and J. Kielman, "Challenges for visual analytics," *Inf. Vis.*, vol. 8, no. 4, pp. 309–314, 2009.

[79] M. Riveiro and G. Falkman, "Interactive visualization of normal behavioral models and expert rules for maritime anomaly detection," in *Proc. 6th Int. Conf. Comput. Graphics Imaging Vis.*, 2009, pp. 459–466.

[80] S. Foresti, J. Agutter, Y. Livnat, S. Moon, and R. F. Erbacher, "Visual correlation of network alerts," *IEEE Comput. Graphics Appl.*, vol. 26, no. 2, pp. 48–59, Mar./Apr. 2006.

[81] S. T. Teoh, K. Zhang, S.-M. Tseng, K.-L. Ma, and S. F. Wu, "Combining visual and automated data mining for near-real-time anomaly detection and analysis in BGP," in *Proc. ACM Workshop Vis. Data Mining Comput. Security*, 2004, pp. 35–44.

[82] X. Yin, W. Yurcik, M. Treaster, Y. Li, and K. Lakkaraju, "Visflowconnect: Netflow visualizations of link relationships for security situational awareness," in *Proc. ACM Workshop Vis. Data Mining Comput. Security*, 2004, pp. 26–34.

[83] T. Taylor, D. Paterson, J. Glanfield, C. Gates, S. Brooks, and J. McHugh, "Flovis: Flow visualization system," in *Proc. Cybersecurity Appl. Technol. Conf. Homeland Security*, 2009, pp. 186–198.

[84] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi, "Preserving the big picture: Visual network traffic analysis with tnv," in *Proc. IEEE Workshop Vis. Comput. Security*, 2005, pp. 47–54.

[85] S. T. Teoh, K.-L. Ma, S. F. Wu, and T. J. Jankun-Kelly, "Detecting flaws and intruders with visual data analysis," *IEEE Comput. Graphics Appl.*, vol. 24, no. 5, pp. 27–35, Sep./Oct. 2004.

[86] F. Fischer and D. A. Keim, "Nstreamaware: Real-time visual analytics for data streams to enhance situational awareness," in *Proc. 11th Workshop Vis. Cyber Security*, 2014, pp. 65–72.

[87] J. Tao *et al.*, "Visual analysis of collective anomalies through high-order correlation graph," in *Proc. IEEE Pacific Vis. Symp.*, 2018, pp. 150–159.

[88] E. Bertini, P. Hertzog, and D. Lalanne, "Spiralview: Towards security policies assessment through visual correlation of network resources with evolution of alarms," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, 2007, pp. 139–146.

[89] S. T. Teoh, K. L. Ma, S. F. Wu, and X. Zhao, "Case study: Interactive visualization for internet security," in *Proc. IEEE Conf. Vis.*, 2002, pp. 505–508.

[90] Y. Shi, Y. Zhao, F. Zhou, R. Shi, and Y. Zhang, "A novel radial visualization of intrusion detection alerts," *IEEE Comput. Graphics Appl.*, vol. 38, no. 6, pp. 83–95, Nov./Dec. 2018.

[91] J. R. Goodall *et al.*, "Situ: Identifying and explaining suspicious behavior in networks," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 204–214, Jan. 2019.

[92] Y. Livnat, J. Agutter, S. Moon, R. F. Erbacher, and S. Foresti, "A visualization paradigm for network intrusion detection," in *Proc. 6th Annu. IEEE SMC Inf. Assurance Workshop*, 2005, pp. 92–99.

[93] F. Mansmann, D. A. Keim, S. C. North, B. Rexroad, and D. Sheleheda, "Visual analysis of network traffic: interactive monitoring, detection, and interpretation of security threats," *IEEE Trans. Vis. Comput. Graphics*, vol. 13, no. 6, pp. 1105–1112, 2007.

[94] R. F. Erbacher, K. L. Walker, and D. A. Frincke, "Intrusion and misuse detection in large-scale systems," *IEEE Comput. Graphics Appl.*, vol. 22, no. 1, pp. 38–47, Jan./Feb. 2002.

[95] K. Lakkaraju, W. Yurcik, and A. J. Lee, "NVisionIP: Netflow visualizations of system state for security situational awareness," in *Proc. ACM Workshop Vis. Data Mining Comput. Security*, 2004, pp. 65–72.

[96] C. Xie, W. Xu, and K. Mueller, "A visual analytics framework for the detection of anomalous call stack trees in high performance computing applications," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 215–224, Jan. 2019.

[97] A. Boschetti, L. Salgarelli, C. Muelder, and K.-L. Ma, "TVi: A visual querying system for network monitoring and anomaly detection," in *Proc. 8th Int. Symp. Vis. Cyber Security*, 2011, Art. no. 1.

[98] F. Mansmann and S. Vinnik, "Interactive exploration of data traffic with hierarchical network maps," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 6, pp. 1440–1449, Nov./Dec. 2006.

[99] A. D. D'Amico, J. R. Goodall, D. R. Tesone, and J. K. Kopylec, "Visual discovery in computer network defense," *IEEE Comput. Graphics Appl.*, vol. 27, no. 5, pp. 20–27, Sep./Oct. 2007.

[100] B. Mukherjee, L. T. Heberlein, and K. L. Levitt, "Network intrusion detection," *IEEE Netw.*, vol. 8, no. 3, pp. 26–41, May/Jun. 1994.

[101] M. L. Huang, J. Liang, and Q. V. Nguyen, "A visualization approach for frauds detection in financial market," in *Proc. 13th Int. Conf. Inf. Vis.*, 2009, pp. 197–202.

[102] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Knowl.-Based Syst.*, vol. 70, pp. 324–334, 2014.

[103] M. C. Hao, D. A. Keim, U. Dayal, and J. Schneidewind, "Visimpact: Business impact visualization," in *Visualization and Data Analysis 2005*, vol. 5669. Bellingham, WA, USA: International Society for Optics and Photonics, 2005, pp. 238–250.

[104] M. Suntinger, H. Obweger, J. Schiefer, and M. E. Groller, "The event tunnel: Interactive visualization of complex event streams for business process pattern analysis," in *Proc. IEEE Pacific Vis. Symp.*, 2008, pp. 111–118.

[105] R. A. Leite *et al.*, "EVA: Visual analytics to identify fraudulent events," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 1, pp. 330–339, Jan. 2018.

[106] M. C. Hao, D. A. Keim, and U. Dayal, "Visbiz: A simplified visualization of business operation," in *Proc. Conf. Vis.*, 2004, pp. 598–1.

[107] Z. Niu, D. Cheng, L. Zhang, and J. Zhang, "Visual analytics for networked-guarantee loans risk management," in *Proc. IEEE Pacific Vis. Symp.*, 2018, pp. 160–169.

[108] R. A. Leite, T. Gschwandtner, S. Miksch, E. Gstrein, and J. Kuntner, "Visual analytics for fraud detection: Focusing on profile analysis," in *Proc. Eurographics/IEEE VGTC Conf. Vis.: Posters*, 2016, pp. 45–47.

[109] M. C. Hao, D. A. Keim, U. Dayal, and J. Schneidewind, "Business process impact visualization and anomaly detection," *Inf. Vis.*, vol. 5, no. 1, pp. 15–27, 2006.

[110] P. A. Legg, "Visualizing the insider threat: challenges and tools for identifying malicious user activity," in *Proc. IEEE Symp. Vis. Cyber Security*, 2015, pp. 1–7.

[111] W. Didimo, G. Liotta, F. Montecchiani, and P. Palladino, "An advanced network visualization system for financial crime detection," in *Proc. IEEE Pacific Vis. Symp.*, 2011, pp. 203–210.

[112] R. Chang *et al.*, "Wirevis: Visualization of categorical, time-varying data from financial transactions," in *Proc. IEEE Symp. Vis. Analytics Sci. Technol.*, 2007, pp. 155–162.

[113] R. Chang *et al.*, "Scalable and interactive visual analysis of financial wire transactions for fraud detection," *Inf. Vis.*, vol. 7, no. 1, pp. 63–76, 2008.

[114] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko, "Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 10, pp. 1646–1663, Oct. 2013.

[115] E. N. Argyriou, A. Symvonis, and V. Vassiliou, "A fraud detection visualization system utilizing radial drawings and heatmaps," in *Proc. Int. Conf. Inf. Vis. Theory Appl.*, 2014, pp. 153–160.

[116] E. N. Argyriou, A. A. Sotiraki, and A. Symvonis, "Occupational fraud detection through visualization," in *Proc. IEEE Int. Conf. Intell. Security Inform.*, 2013, pp. 4–6.
[117] Y.-A. Kang and J. Stasko, "Examining the use of a visual analytics system for sensemaking tasks: Case studies with domain experts," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 12, pp. 2869–2878, Dec. 2012.
[118] W. Didimo, L. Giamminonni, G. Liotta, F. Montecchiani, and D. Pagliuca, "A visual analytics system to support tax evasion discovery," *Decis. Support Syst.*, vol. 110, pp. 71–83, 2018.
[119] J. A. Guerra-Gomez, A. Wilson, J. Liu, D. Davies, P. Jarvis, and E. Bier, "Network explorer: Design, implementation, and real world deployment of a large network visualization tool," in *Proc. Int. Working Conf. Advanced Vis. Interfaces*, 2016, pp. 108–111.
[120] Y. Sun, Y. Tao, G. Yang, and H. Lin, "Visitpedia: Wiki article visit log visualization for event exploration," in *Proc. Int. Conf. Comput.-Aided Design Comput. Graphics*, 2013, pp. 282–289.
[121] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proc. IEEE Symp. Vis. Languages*, 1996, pp. 336–343.
[122] M. E. Joorabchi, J.-D. Yim, and C. D. Shaw, "Emailtime: Visual analytics of emails," in *Proc. IEEE Symp. Vis. Anal. Sci. Technol.*, 2010, pp. 233–234.

**Yang Shi** received the PhD degree in computer science from Central South University, China, in 2017. She is currently an assistant researcher with the College of Design and Innovation of Tongji University, Shanghai, China. Her current research interests include data visualization and human computer interaction.

**Yuyin Liu** received the BSc degree in physics with Honours from Imperial College London, U.K, in 2018. She is currently working toward the master's degree at the Chinese University of Hong Kong, China.

**Hanghang Tong** received the MSc and PhD degrees in machine learning from Carnegie Mellon University, United States, in 2008 and 2009, respectively. He is currently an associate professor with the University of Illinois at Urbana-Champaign, United States. His research interest includes large scale data mining for graphs and multimedia.

**Jingrui He** received the MSc and PhD degrees in machine learning from Carnegie Mellon University, United States, in 2008 and 2010, respectively. She is currently an associate professor with the University of Illinois at Urbana-Champaign, United States. Her research interests include heterogeneous machine learning, rare category analysis, active learning and semi-supervised learning, with applications in social network analysis, healthcare, manufacturing, etc.

**Gang Yan** received the BSc and PhD degrees from the University of Science and Technology of China, china, in 2005 and 2010, respectively. He is currently a professor with the School of Physics Science and Engineering, Tongji University, Shanghai, China. His research interests include network science, biological and artificial neural networks.

**Nan Cao** is currently a professor at Tongji University, and the director of Intelligent Big Data Visualization Lab (iDV$^{x}$ Lab). His research interests include data visualization, visual analysis, and data mining. He creates novel visual analysis techniques for supporting anomaly detection in complex (i.e., multivariate, heterogeneous, and multi-relational) data.