TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems

Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, Ching-Yung Lin

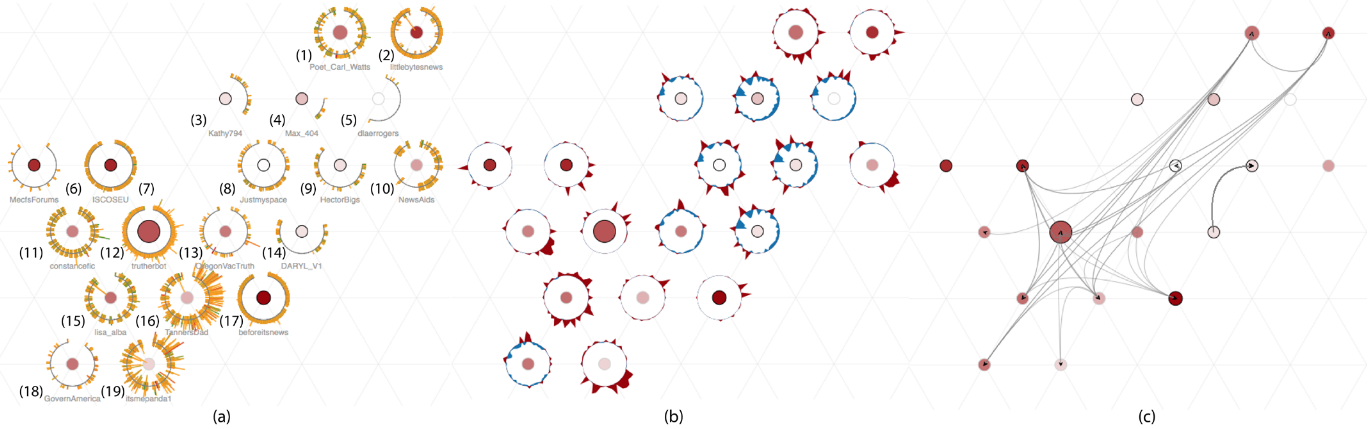


Fig. 1. The visualization of top ranking anomalous twitter users. In this visualization, users, labeled by (1 - 19), are represented as circles sized by their importances (i.e., number of followers), colored by their anomaly scores ranging from white (lowest) to dark red (highest). In different visualization modes, each circle is surrounded by a visualization of (a) the user's activity threads, or (b) z-scores of the user's features, or (c) links indicating the user's interactions with others.

Abstract—Users with anomalous behaviors in online communication systems (e.g. email and social medial platforms) are potential threats to society. Automated anomaly detection based on advanced machine learning techniques has been developed to combat this issue; challenges remain, though, due to the difficulty of obtaining proper ground truth for model training and evaluation. Therefore, substantial human judgment on the automated analysis results is often required to better adjust the performance of anomaly detection. Unfortunately, techniques that allow users to understand the analysis results more efficiently, to make a confident judgment about anomalies, and to explore data in their context, are still lacking. In this paper, we propose a novel visual analysis system, TargetVue, which detects anomalous users via an unsupervised learning model and visualizes the behaviors of suspicious users in behavior-rich context through novel visualization designs and multiple coordinated contextual views. Particularly, TargetVue incorporates three new ego-centric glyphs to visually summarize a user's behaviors which effectively present the user's communication activities, features, and social interactions. An efficient layout method is proposed to place these glyphs on a triangle grid, which captures similarities among users and facilitates comparisons of behaviors of different users. We demonstrate the power of TargetVue through its application in a social bot detection challenge using Twitter data, a case study based on email records, and an interview with expert users. Our evaluation shows that TargetVue is beneficial to the detection of users with anomalous communication behaviors.

Index Terms—Anomaly Detection, Social Media, Visual Analysis

1 Introduction

Recently, online communication systems, such as email and social media platforms (e.g., Twitter and Facebook), provide new mechanisms for users to share information with each other across space and time. Everyday, these systems generate enormous digital data archives recording various user activities, introducing a proliferation of opportunities to understand users' communication behaviors. Analyzing these behaviors not only helps with finding the common communication patterns adopted by the public, but more importantly facilitates the detection of anomalous users who are potential threats to society. The problem of anomaly detection [8] has attracted great attentions in the field of machine learning. Many analysis methods, both supervised [33] and unsupervised [11], have been developed to address this problem. All of the anomaly detection methods face the challenge that the ground truth required for training and/or for performance evaluation is usually difficult to obtain. Manually annotating data, a common approach to address this issue, is mostly tedious, time consuming, and fully dependent on the judgment of the annotators, which can greatly impact the quality of the analysis results.

Data visualization provides a great means to evaluate the analysis results via intuitive representations of context information that provides additional evidences to support or refute the analysis conclusions. However, to design an effective visualization to portray users' behavioral patterns in a communication process, three challenges need to be tackled: (1) display and capture the rich contexts of a communication process through a simple and integrated visual design to facilitate efficient visual comparison; (2) capture how the temporal patterns (e.g., frequency and duration of the communication process), content patterns (e.g., the topics around which the interaction occurred), and activity patterns (e.g., how a user posts on Twitter) are important for revealing the insight of a user's behavior; (3) design a generalized visualization to support anomaly detection of users based on various

- *Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, and Ching-Yung Lin are with IBM T. J. Watson Research Center. E-mail: {nancao, conglei, sabrinal, jielv, chingyung}@us.ibm.com.*
- *Yu-Ru Lin is with University of Pissburg. E-mail: yurulin@pitt.edu.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467196

data collected from different communication systems, where a standard approach or common understanding of the underlying structures of a typical communication process is lacking.

To address these challenges, we introduce TargetVue, a novel visual analysis system for detecting, summarizing, interpreting, and comparing anomalous user behaviors archived in various types of communication data. TargetVue employs an unsupervised learning model, TLOF [2] (a well-studied anomaly detection technique), to detect and rank anomalous users based on a set of well-defined features. Multiple coordinated views are employed in TargetVue to visually summarize and represent the analysis results as well as various important aspects of the users' communication behaviors. These aspects include topics, sentiments, temporal dynamics of the users' communication features and their impacts, as well as the relationships among different users. The coordinated view allows analysts to browse and compare users' communication behaviors in TargetVue from different perspectives. Specifically, the contributions of this paper include:

- **System.** We introduce a novel visual analysis system leveraging advanced machine learning algorithms and visualization techniques to detect and support interactive exploration of anomalous users with various visual representations and view perspectives. We also identify several high-level feature types, through an anatomy of a typical social communication procedure, to apply the system to different communication data.

- **Visualization.** We propose new glyph designs and layout algorithm for efficiently summarizing and comparing different communication behaviors. Particularly, we introduce three types of glyphs, the activity glyph, z-glyph, and relation glyph, to capture individual user's behaviors based on their communication activities (i.e., posting and responding) and corresponding features, as well as their interactions with others. A global layout algorithm is also developed for efficiently positioning users based on their similarities in a triangle grid, facilitating the visual comparison and clustering of their behavior glyphs.

- **Evaluation.** We demonstrate the power of the TargetVue system in a bot detection challenge on Twitter and also conducted a case study based on an email dataset. We highlight several interesting findings as well as visual patterns of the anomalous user behaviors based on our visual designs.

The rest of this paper is organized as follows. We first discuss related work in Section 2, followed by system overview and data processing pipeline in Section 3. Section 4 introduces the model and features used by TargetVue for anomaly detection. We present detailed design requirements, rationales, and techniques in Section 5. In Section 6, we describe a comprehensive evaluation of the proposed system, including its application in a Twitter bot detection challenge, a case study using email data, and expert interviews. Finally, in Section 7, we conclude with a summary and future directions.

2 Related Work

In this section, we survey the papers that are most related to our work, including anomaly detection, visual analysis of user behaviors, and visual summarization of activities.

2.1 Anomaly Detection

Given its broad impact on security systems, anomaly detection has been extensively studied over the past decades and a wide variety of anomaly detection methodologies have been proposed [8]. One category of anomaly detection methods employs supervised learning approaches by training models for both normal and anomalous classes based on labeled training data [33]. Another category applies unsupervised learning which identifies anomalies either assuming most of the training data as normal or requiring no training data [11]. All such anomaly detection approaches face the issue of the lack of the ground truth, making evaluation difficult.

More and more visualization techniques have been applied to help with anomaly detection and evaluation. Particularly, statistical diagrams such as time series charts and histograms are most commonly used to represent the anomalous changes in the raw data [19, 23, 25]. Various types of dimension reduction and multidimensional visualizations techniques such as multidimensional scaling (MDS) [21], principle component analysis (PCA) [16], self-organization map (SOM) [20], and parallel coordinates [14] are also used to represent the data's distributions in a multidimensional feature space, thus facilitating outlier detection [17, 26, 27]. Particularly, there have been substantial works focusing on visualizing the traffic data of computer networks for intrusion detection [1, 9, 34, 37]. However, these works, given their narrow focus, can hardly be applied to other applications.

Most recently, some researchers investigate anomalies in social media data, which are more relevant to our work. Particularly, Thom *et al.* [35] introduced a visual analysis system for monitoring anomalous bursting of keywords at different times and locations based on a tag cloud visualization overlaid on top of a map. Zhao *et al.* [45] developed the FluxFlow system for detecting and visualizing anomalous information propagation processes in Twitter. This system employs the OCCRF model [32] and interprets the analysis results in multiple visualization views, showing different context information about the propagation such as topics, sentiments, features, and involving users. Compared to these systems, TargetVue focuses on detecting and visualizing another type of anomaly, i.e., anomalous user behaviors. This is a more fundamental problem as the users' behaviors determined how messages were posted and spread. Moreover, TargetVue is more broadly designed to deal with the data archived from all types of online communication systems and supports different visual designs and components. In addition, TargetVue adopts a more efficient anomaly detection model, TLOF [2], whose results are easier to be interpreted in visualizations.

2.2 Visual Analysis of User Behaviors

In recent years, a great number of visualizations have been developed to represent email records or social media data [3, 5, 6, 45]. Here, we only focus on relevant designs that are user centric, i.e., the ones that analyze the features and behaviors of individuals or groups of users. There have been several papers about modeling users' behaviors but little work has been done on visualizing them. Most existing visualizations in this topic are designed based on email data. For example, Li *et al.* [24] visualized an individual's email communication flows and cliques of organizational email accounts for detecting anomaly based on clique violation. Perer *et al.* [29] summarized and compared the trends of email activities based on the numbers of messages over the duration of an email archive for relationship discovery. Viégas *et al.* [39] presented a visualization showing detailed email exchange history between the user and friends for relationship characterization. Different from all these designs, in TargetVue, we summarize a user's communication behaviors in glyphs, which are simple and clean, thus facilitating an efficient comparison over multiple users.

Many analysis-driven approaches were introduced for understanding user behaviors. Kumar *et al.* [22] studied users' migration behaviors among different social media platforms and represented the findings via radar charts. Ratkiewicz *et al.* [30] investigated one special type of malicious user behaviors, "astroturf", in Twitter during political campaigns. They identified these behaviors by closely monitoring users' mentioning of a set of pre-defined social memes in their tweets and conducted classification to separate different user behaviors. They also visualized the number of mentions of social memes in time-series charts and represented the spreading of these memes among users via a node-link graph. Pennacchiotti *et al.* [28] introduced a classification framework of Twitter users based on features about their tweeting behaviors and topics. Tinati *et al.* [36] proposed techniques for identifying communicators' roles in Twitter. They named a set of roles including "idea starter", "amplifier", "curator", and "commentator", defined statistical models for role identification, and visualized a social network with node colors indicating roles. Viswanath *et al.* [41] studied the evolution of an interaction graph based on Facebook data to aid the understanding of how users communicate with each other over time. Java *et al.* [15] explored user communities in Twitter based on users' retweeting behaviors. Cha *et al.* [7] introduced a formal measurements of a user's influence in Twitter. Goncalves et.al [13] applied a physical model on Twitter networks to capture users' activities. Yang *et al.* [44] and Xu *et al.* [43] introduced models for capturing users' retweeting and posting behaviors respectively.

The statistical analyses of user behaviors or activities from existing work provide necessary theoretical supports for our work and also inspired many of our designs. Their limitations such as lack of context, difficulty of understanding and evaluation, and improper use of visualization tools, motivated us to develop an advanced visualization for illustrating user behaviors and facilitating the procedure of anomaly detection.

2.3 Visual Summarization of Activities

Some visualizations have been designed for summarizing different types of activities. Novel glyph designs are introduced to produce a highly identifiable representation of various activities. For example Erbacher *et al.* [10] introduced a radial glyph that shows a web server's activities for connecting to other servers over time. Fry [12] introduced a glyph that statistically shows users' visits to a web page. These designs summarized the activities at a given time point as a glyph and the changes of activities were displayed frame-by-frame. Xiong et.al [42] developed PeopleGarden, a flower shaped glyph, for summarizing a user's aggregated interaction histories in a discussion group. The flower glyphs of different users are randomly positioned in a display area called "garden". Although it summarized users' interactions, all the details such as "when did who get involved in a communication procedure" are unavailable from the visualization. These designs may be useful in providing a snapshot view or an aggregated view of users' behaviors, but they are not effective in identifying or comparing temporal patterns from the data. Viegas et.al [40] introduced HistoryFlow, a stacked flow visualization that displays collaborations of the users who edited on the same Wikipedia page. This visualization allowed users to compare interaction (i.e. co-editing a page) patterns with respect to a limited interaction context (a single Wiki page). It is thus difficult to be extended to visualize or compare the change of communication contexts over time.

3 System Design and Overview

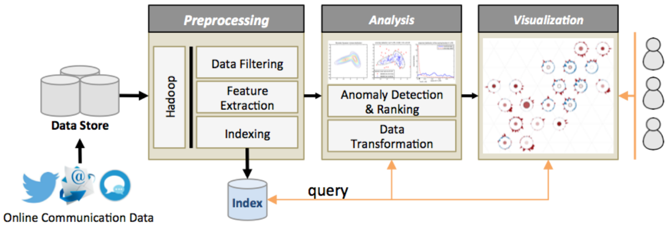


Fig. 2. System overview and data processing pipeline.

The TargetVue system was part of a four-year anomaly detection project initiated by a major defense agency. This project was aimed at satisfying several real-world requirements for anomaly detection and was supervised by two domain experts – a counter-intelligence analyst and a research director whose team focused on uncovering malicious user behaviors in, e.g., enterprises, online social media, and email systems. Regular research discussion meetings with these experts and review meetings with the project sponsor were held. During these meetings, the experts and the review board clarified their requirements and evaluated the prototypes developed for the project. They also provided many constructive suggestions to improve the system. Below we list the requirements that are most relevant to the design of TargetVue:

R1 Feature Selection. Determine what are the right features for detecting malicious user behaviors corresponding to any given social communication data.

R2 Anomaly Detection in Contexts. Incorporate auxiliary information from the data to form a semantic background against which data will be interpreted and apply the semantic layer to anomalies detected to drive down false positive rates to levels manageable by a human operator.

R3 Ranking Threats. The software will take resulting anomalous behaviors and rank them in importance (for the benefit of a human operator) regarding their potential as a significant emergent threat. Semantics will be used in the threat ranking process. Rankings will include justifications based on semantics.

R4 Learn from User Feedback. Operators will use justifications as a basis for critiquing rankings. Critiques will be fed back to improve the anomaly detection procedures.

Based on these requirements, we have developed the TargetVue system. Fig. 2 illustrates the system architecture and the data processing pipeline. The system consists of four modules: (1) the data collection module, (2) the preprocessing module, (3) the analysis module, and (4) the visualization module. The data collection module collects and stores the online social communication data such as tweets, emails, and instant messages in an offline procedure. These data are subsequently processed in the preprocessing module that runs on a cluster based on Apache Hadoop. At this stage, the system conducts data filtering to discover the high impact users who actively posted or replied to a large amount of messages during the communication. Key features are extracted for each user guided by a tripartite graph (described in Section 4.2) that models the essential components in a typical social communication process (**R1**). A full-text index of the messages is also built, facilitating a topic-driven data exploration (**R2**). The analysis module runs anomaly detection algorithms to detect users with suspicious communication behaviors based on their features and ranks these users based on their anomaly scores (**R3**). The visualization module displays anomalous users together with the corresponding contexts and raw data records within several views, providing a comprehensive visual summarization and interpretation of users' communication behaviors (**R2**), thus facilitating annotation of the data and evaluation of analysis results (**R4**).

All these modules work together to form a scalable mechanism that enables an efficient procedure to reduce the information seeking space. Particularly, powered by Hadoop, the data preprocessing module is able to deal with millions of messages in an offline procedure within hours; by employing a fast anomaly detection algorithm, the analysis module is able to rank tens of thousands of users in near real-time; the visualization module introduces views to further represent data at various finer granularities. For example, a global view is designed to illustrate several hundreds of top-ranked suspicious users and an investigation view and several other contextual views are designed to illustrate the detailed behaviors of a few dozens of focused users.

4 Detecting Anomalous User Behaviors

In this section, we introduce the model and high-level features used by TargetVue for detecting anomalies in online communication systems.

4.1 Time-Adaptive Local Outliner Factor

One challenge in anomaly detection is that there may be no labeled anomalies or the anomalies are too few to accurately represent the underlying distribution of the anomaly class. It is even more challenging in online communications when the environment is highly dynamic and the anomalies should be detected quickly with as little training data as possible. With these considerations in mind, we adapted the time-adaptive local outlier factor model (TLOF) [2] to identify anomalies as the sudden changes of user behaviors based on a set of features extracted for each user from the online communication data. We choose this model because of many of its advantages: (1) it is an unsupervised learning model requiring no training data, which fits our application scenario, i.e., no anomalous users are known in advance; (2) it takes the time sequence of user behavior into account instead of just one snapshot of the behavior to reduce false positives; (3) it assigns anomaly scores instead of binary labels (normal or anomaly) to users, thus providing a ranked list of users which is important for reducing the searching space during the visual analysis procedure; (4) it detects outliers based on Euclidean distance, making the results easily interpretable by visualizations. Finally, this algorithm is efficient and is able to compute results in near real-time (the time complexity is $O(NlogN)$, where N is the number of users), thus allowing it to be easily integrated into the interactive visual analysis system.

Formally, we describe a user's behaviors by a time series of feature vectors, $X = [x_1, x_2, ..., x_T]$, where x_t is a feature vector describing the user's behaviors observed at time $t \in [1, 2, ..., T]$. The TLOF gives an anomaly measurement for every time series (i.e. every user) by identifying the features that are significantly different from other series in the test data and the past history of its own. Formally, a user's anomaly score, $s(X)$, is defined as follows:

$$s(X) = \alpha \cdot Z_1(X) + (1 - \alpha) \cdot Z_2(X) \quad (1)$$

$$Z_1(X) = LOF(x_T) - \sum_{t=T-W}^{t=T-1} LOF(x_t)/W$$

$$Z_2(X) = 1 - P_N(LOF(x_T), \mu, \sigma)$$

where $LOF(x_t)$ is the local outlier factor of the behavior feature vector x_t, which is based on x_t's neighborhood density:

$$LOF_k(x_t) = \frac{\sum_{y \in N_k(x_t)} D_k(y_t)}{|N_k(x_t)| D_k(x_t)} \quad (2)$$

$$D_k(x_t) = \frac{|N_k(x_t)|}{\sum_{q \in N_k(x_t)} (max(d_k(q), d(x_t, q)))}$$

where $N_k(x_t)$ is the set of the k-nearest neighbors of x_t in the feature space; $D_k(x_t)$ is x_t's neighborhood density; $d(x_t, q)$ is the Euclidean distance between x_t and its neighbor q; $d_k(q)$ is the maximum distance between q and its k-nearest neighbors. Intuitively, $LOF(x_t)$ is designed based on the assumption that users' behaviors form several latent clusters in the feature space, hence x_t's outlier factor can be determined by only comparing it to the nearby feature vectors in the feature space instead of the entire population.

In equation (1), the weight $\alpha \in [0, 1]$ balances between two terms. The first term (Z_1) indicates the difference between the current LOF value and the average LOF value in the past within the time window W. The second term (Z_2) estimates the probability of x_t being considered as an outlier under the normal distribution $P_N(\cdot)$ with mean μ and standard deviation σ that are computed over the period of the whole time sequence X. A higher value of $Z_2(x_T)$ or a lower value of $P_N(x_T)$ indicates a higher probability of the behavior described by x_T being considered to be an anomaly.

4.2 Communication Features

In addition to the analysis model, extracting a set of feasible features for describing different social behaviors of users is another important task for building the system (**R2**). However, this is not easy since it not only requires a deep understanding of the users' behaviors in different social communication platforms, but also requires extracting a set of common features to capture the essential characteristics of the problem across different types of social communication data. For this purpose, we investigated different social communication processes and decomposed them into essential components that are structured in a generalized data model [3].

Specifically, social communications usually involve *social objects*, i.e., the content around which a conversation happens [31] such as emails (in email exchanges), tweets (in Twitter communications), and messages (in instant messaging). A social object connects people with shared interests in a social communication. These people usually play in two types of roles in a communication procedure: an *initiator* who initiates the interaction by creating a social object, and a *responder* who responds by acting on the social object created by the initiator. These concepts can be captured in a tripartite graph model as shown in Fig. 3, in which initiators, responders and social objects are differentiated as three different types of nodes. Both initiating and responding actions with respect to social objects are denoted as directed links that are labeled by timestamps, showing the time when the corresponding actions occurred.

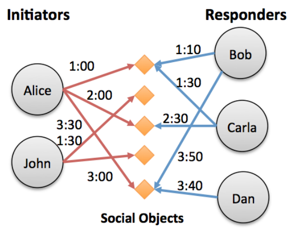


Fig. 3. Data model for social communication.

Based on this data model, we identify six high-level feature categories for capturing users' communication behaviors by considering their roles, interactions, and the corresponding social objects.

F1 Behavior Features. Features in this category aim to identify users' roles based on their posting or responding behaviors. We believe that users play in different roles may behave abnormally in different ways. For example, in Twitter, a "spammer" may behave as an initiator most of the time, but an "information spreader" who closely monitors certain topics may retweet a lot, acting as a responder most of the time. Thus, classifying users based on their roles can help interpret their behaviors correctly in diverse contexts.

F2 Content Features. Features in this category, such as topical keywords, sentiment scores or the amount of special tags or symbols, focus on the properties of social objects. We assume a sudden change of these features may imply an anomalous behavior, for example, talking about a sensitive topic (sudden change of topics) or maliciously attacking another user with negative sentiments in tweet text.

F3 Interaction Features. These features focus on describing how users communicate with others and how others respond to them. Do they tend to communicate with a small group of users or broadcast messages to the public? During the communication, do they have a decent conversation reciprocity? For example, a spammer may send spams to a variety of users but they will not be responded to in most cases.

F4 Temporal Features. Features in this category, such as posting/replying/receiving interval/frequency entropy, measure the regularity of certain types of user behaviors. Our hypothesis is that temporal features of normal users are more or less random over time. Thus regularities in these features are suspicious.

F5 Network Features. Features such as users' in/out degrees in a network provide ego-centric measurements of the social network structure in different aspects. A sudden change of these features may imply anomalous events or behaviors.

F6 User Profile Features. Sometimes user profiles can also reveal anomalous behaviors. For example, in Twitter, a social bot may frequently change its screen name to pretend to be others.

These high-level feature categories help us identify useful features from different datasets. Later, in Section 6, we will show how they help us in detecting anomalous users from Twitter and email data.

5 User Interface and Visualization

In this section, we describe the design tasks derived from the discussions with our expert users, followed by the detailed description of the visualization views whose designs were driven by these tasks.

5.1 Design Tasks

In addition to the general requirements mentioned in Section 3, we have discussed with the experts about difficulties they encounter when trying to decide if a user is anomalous. The most commonly mentioned difficulty is that the size of the raw data is too large for the analysts to go through every single user and activities even when the users are presented in a ranking list. They desire a tool that can help them to efficiently review the results reported by the anomaly detection algorithm in order to quickly identify false positives. In addition, the experts seek for a tool that allow them to label the results. To meet these requirements, we decided on a list of visualization design tasks as follows.

T1 Showing the data overview and detection results. The number of users as well as the number of activities in online communication are in the scale of tens of thousands or more. Hence for each user, the analysts need one simple visualization that can summarize the activities and the anomaly detection results so that they can quickly skim through the whole population.

T2 Interpreting user behaviors from different perspectives. Having a comprehensive understanding of the semantics of user behaviors is important for anomaly detection and eliminating false positives. Thus the visualization should be able to present users in a full range of contexts including their communication topics, activities (posting/responding), features, and relationships between users (both implicit relationships such as similarity in the feature space and explicit relationships such as mutual communication or friendships).

T3 Facilitating visual data comparisons. Another key to understanding the patterns of different user behaviors is the ability to differentiate patterns, especially to differentiate anomalous users from normal ones. Hence, the system should facilitate pattern comparison through symbolic representation of behaviors and interactions.

T4 Revealing users' impacts in social communications. This helps the analysts to estimate the potential threats from a suspicious user as required in **R3**. Users' impacts can be determined by their profile features such as the number of contacts or responders, but more importantly, they can be dynamically estimated from the propagation patterns of the messages which the users are involved with [18]. Therefore, visually summarizing both the users' properties and the information propagation history of the messages posted or responded to by the users is also important for visualization and UI design.

T5 Easy browsing of raw data. The raw data such as text written in each message is the strongest support for determining if a user is anomalous of interests. The visualization should enable analysts to explore the raw data easily.

T6 Flexible data labeling. Finally, as stated in **R4**, we should design the UI, visualizations, and corresponding interactions to support data labeling functionality for collecting feedback to the anomaly detection models.

5.2 User Interface

The aforementioned tasks guided our designs of the user interface. As shown in Fig. 4, the UI consists of six major views (**T1, T2**), including (1) the global view displaying the distributions of all users in the feature space; (2) the user list representing detailed profile information of the users; (3) the message list showing the raw communication records of the user currently in focus; (4) the inspection view for visualizing and comparing user behaviors through different glyph representations; (5) the feature variation view showing the changes of users' feature values over time; and (6) the propagation view illustrating the users' impact based on message spreading patterns. These views are interactively connected, illustrating different contexts of a set of top suspicious users ranked by the TLOF model.

We employ consistent visual designs and color coding schemes in all the visualization views: users are visualized as circular nodes (except (6)) sized by their importance (e.g., the number of followers on Twitter) and colored either by their sentiments[1] or anomaly scores. Two sets of color coding schemes have been used. The colors range from light red, to yellow, and to light green are used for indicating most negative, neutral, and most positive sentiments respectively. Another set of colors ranging from dark blue, to white, and to dark red have been consistently used to encode three different anomaly measurements (i.e., TLOF anomaly score, degree of outlierness in the global view, and z-score of features), ranging from -1 to 1. Here, both -1 (dark blue) and 1 (dark red) indicate most anomalous in two opposite directions and 0 (white) indicates normal. More design details of each view are introduced in the following sections.

Use case scenario. To understand how different views work together for a visual analysis task, let us consider the following scenario of using TargetVue's UI to investigate the behaviors of a set of suspicious users. Suppose Alice is a counter-intelligence analyst whose job is to detect malicious user behaviors in Twitter. She uses TargetVue for this purpose. Her data have been processed and analyzed by the TargetVue system. Alice first queries a topic keyword to load a list of top ranking anomalous users who are related to the topic (Fig. 4.4.a). These users are initially displayed in the user list (Fig. 4.2) and the global view (Fig. 4.1). Alice starts by investigating the users' distributions in the feature space via the global view. A small group of users displayed as outliers in the visualization with high anomaly scores attract her attention. She selects them into the inspection view (Fig. 4.4) by filtering, brushing, and direct mouse picking. The selected users are visualized as glyphs summarizing their behaviors and are laid out based on their similarities in a triangle grid. By comparing these glyphs, Alice notices a user who tirelessly posted a lot of messages over time, but those messages were rarely responded to. After inspecting the users' raw communication records in the message list (Fig. 4.3), Alice believes that the user is a spammer. She reports this finding by labeling the user. Those labeled users are highlighted in the UI and stored on the server, which can be later used for tuning and evaluating the underlying anomaly detection model.

5.3 Global View

The global view displays the distribution of all the users in the feature space using multidimensional scaling (MDS) based on users' mean features across the whole time series (**T1**). We render a contour map based on kernel density estimation (KDE) [38] to illustrate a user's degree of outlierness. Intuitively, a user lying in the high density area is considered to be normal as his behavior (measured by the features) is similar to most of other users. In contrast, the users lying in the marginal low density areas are considered to be the outliers (anomalies). Formally, the density at the position x_u, where the user u is placed, is defined as:

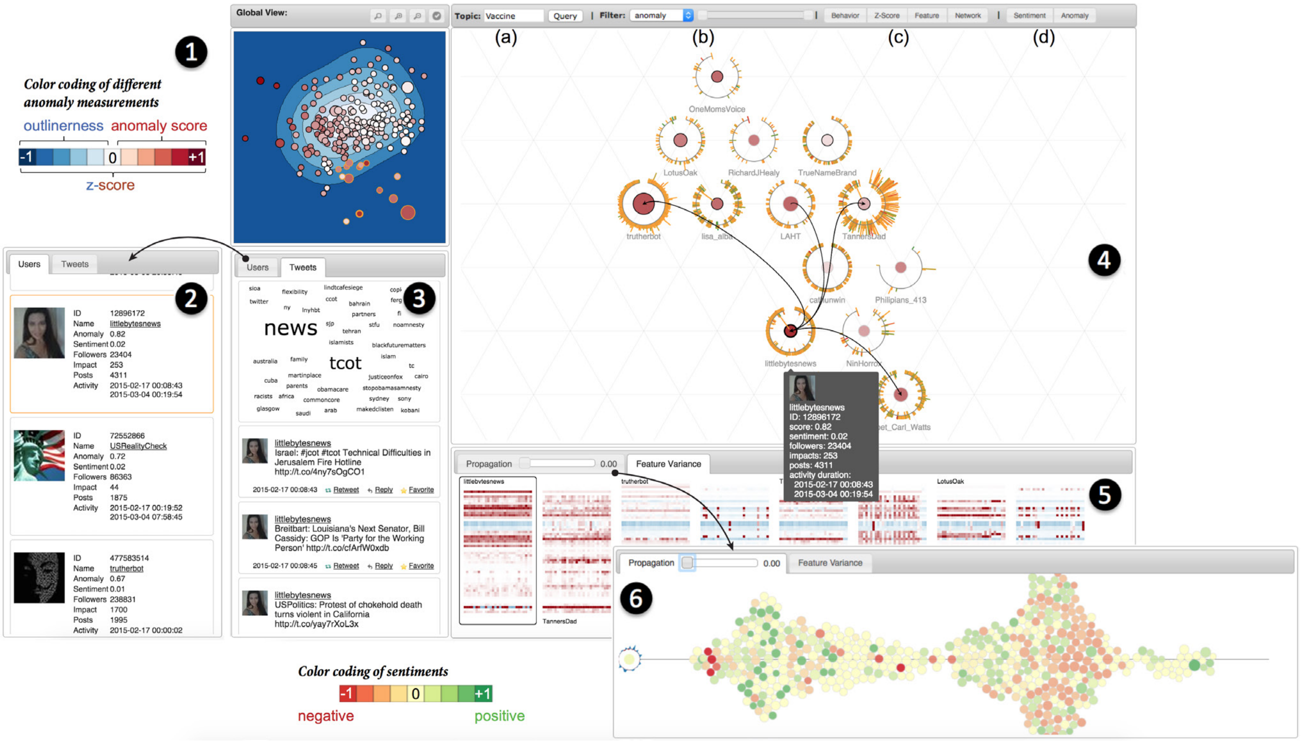
$$f(x_u) = \frac{1}{nh}\sum_{1}^{n} K(\frac{x_u - x_i}{h})$$

where $K(\cdot)$ is the kernel function (e.g., Gaussian kernel), x_i $(i \neq u)$ indicates the positions of other users except u, and h indicates the kernel's bandwidth which is learned in KDE. Contour maps are drawn to reveal areas with different densities colored from white for high density areas (low degree of outlierness) to dark blue for low density areas (high degree of outlierness). Thus, this view not only illustrates an overview of the input data but more importantly provides another type of measurement of anomaly. Inconsistent measurements of the same user or similar users lead to visual cues that may imply interesting patterns. For example, a dark red circle (high TLOF score) shown in the white contour area (low degree of outlierness) may suggest an anomalous user who pretends to be normal, or a normal user who occasionally exhibits anomalous behaviors.

5.4 Inspection View

We design the inspection view to allow a closer look at the details of several individual users chosen from the user list or from the global view. Specifically, as shown in Fig. 1, the users' communication behaviors and the corresponding features are respectively summarized in three types of visual glyphs, *the behavior glyph*, the *z-glyph*, and the *relation glyph*. which are laid out in a triangle grid for easy visual comparison. All these glyphs follow a consistent design in which a

[1] A user's sentiment is the mean value of the sentiments of all his messages.



ID 12896172
Name littlebytesnews
Anomaly 0.82
Sentiment 0.02
Followers 23404
Impact 253
Posts 4311
Activity 2015-02-17 00:08:43
2015-03-04 00:19:54

ID 72552866
Name USRealityCheck
Anomaly 0.72
Sentiment 0.02
Followers 86363
Impact 44
Posts 1875
Activity 2015-02-17 00:19:52
2015-03-04 07:58:45

ID 477583514
Name trutherbot
Anomaly 0.67
Sentiment 0.01
Followers 238831
Impact 1700
Posts 1995
Activity 2015-02-17 00:00:02

littlebytesnews
Israel: #jcot #tcot Technical Difficulties in Jerusalem Fire Hotline http://t.co/4ny7sOgCO1
2015-02-17 00:08:43 Retweet Reply Favorite

littlebytesnews
Breitbart: Louisiana's Next Senator, Bill Cassidy: GOP Is 'Party for the Working Person' http://t.co/cfArfW0xdb
2015-02-17 00:08:45 Retweet Reply Favorite

littlebytesnews
USPolitics: Protest of chokehold death turns violent in California http://t.co/yay7rXoL3x

littlebytesnews
ID: 12896172
score: 0.82
sentiment: 0.02
followers: 23404
impacts: 253
posts: 4311
activity duration:
2015-02-17 00:08:43
2015-03-04 00:19:54

Fig. 4. The user interface of TargetVue system consists of six major views labeled by the numbers 1-6.

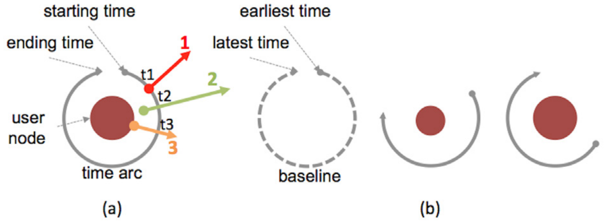


Fig. 5. Behavior glyph design. (a) Design of a single glyph. (b) Visualizing multiple glyphs at the same time.

centric user is represented as a circle in the middle surrounded by a visualization of different types of contexts. This design keeps users' mental map when switching between different views.

Behavior Glyph. This glyph represents two types of activities (i.e., posting and responding) of a centric user using a circular timeline visualization (Fig. 5(a)). This timeline records the history of all the posting and retweeting activities in a clockwise order. Instead of drawing the timeline as a complete circle, a circular arc is used with a gap on top in purpose of avoiding the overlap of starting and ending time, i.e., the time when the first and last activity occurred. Each activity together with the people involved in it, form a thread which is represented as a line segment with length indicating its duration, thickness indicating the number of people involved in the thread (**T4**), and color indicating the sentiment of the corresponding message. Examples of this design are shown as Fig. 5(a-1,2,3). Particularly, the length of a line segment, i.e., the thread duration, is determined by the posting and last retweeting time of the corresponding raw tweet, which are respectively marked by the circular tail and the arrow head of a thread segment. Each thread is perpendicular to the time arc and intersects with it at the time point when the centric user was involved in the thread. For example, when a user posts a message, he is involved from the very beginning, thus the corresponding thread directly connects to the time arc at the time when it is initiated as shown in Fig 5(a-1). When the centric user is involved by responding to a message posted by others, the corresponding thread, as shown in Fig 5(a-2), intersects with the time arc at the responding time t_2. Compared to Fig 5(a-3) which also shows a responding activity, the centric user responded to thread-2 at an earlier stage after it was initiated. The behavior glyph is designed to summarize communication activities and clearly differentiate different types of the activities visually, where information about absolute activity timing is not as important and thus omitted from encoding.

When multiple users are visualized simultaneously in the inspection view, a full range time arc is created as a baseline, indicating the time range between the earliest and the latest time in the data. Thus users' time arcs are drawn in proportion to this range (Fig 5(b)), making them comparable (**T3**). For example, Fig 5(b) illustrates a user's activity started relatively early (middle) or a while late (right).

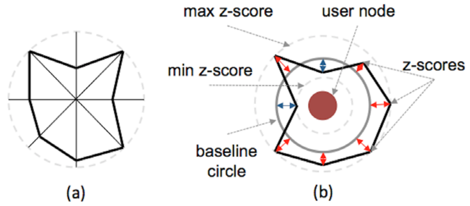


Fig. 6. Feature glyph design. (a) star-glyph; (b) z-glyph showing the differences in the feature values compared to the baseline circle which indicates the mean values of the features.

Z-Glyph. We summarize a user's mean communication features across time in a glyph that was inspired by star glyph and specifically designed for anomaly detection tasks. A star glyph (Fig. 6(a)) represents data features as axises that are radially arranged and share the same starting point at the center of a circle. A data item's feature values are thus plotted on each axis and connected together in a polyline forming a star-shape. Although widely used, this design is ineffective for anomalous detection as no context about the normal condition is shown.

To address this shortcoming, we introduce "z-glyph" to illustrate the differences between the centric user and a normal baseline based on the features' z-scores. Specifically, we compute the baseline feature values as the mean feature values of all the users in the data. The rationale behind this is the one-class nature of the anomaly detection problem: most of the users are considered to be normal and are classified into one big class, and there are only very few users considered as anomalies. Thus, the mean feature values are able to represent the overall behaviors of normal users. We estimate a user's differences from the baseline using z-score defined as follows:

$$Z_i(x) = (f_i(x) - \mu_i)/\sigma_i \tag{3}$$

where $f_i(x)$ and $Z_i(x)$ respectively indicate the user x's i-th feature value and the corresponding z-score value; μ_i and σ_i indicate the mean and standard deviation of the i-th feature over the entire dataset.

Visually, we represent the baseline values as a circle, and plot the user's z-scores inside/outside the circle and connect these plots in a polyline. The positive z-scores are plotted outside the circle and negative ones are plotted inside. Colors are also used to visually enhance the differences. Dark red and blue are used to fill those positive and negative regions respectively (Fig. 1(b)). This way, a normal user will be in a shape close to the baseline circle, whereas the anomalous ones will be illustrated in irregular shapes and colors, a visual metaphor of their irregular behaviors.

Relation Glyph. We illustrate a user's interaction relationships with others as outgoing directed links that start from the centric user connecting to the persons whom the users interact with. These interactions could be, for example, following, mentioning, retweeting, or replying in Twitter or sending and replying to emails. The links are bundled together based on their trend as shown in Fig. 1(c). Multiple relation glyphs together essentially represent a social network.

Layout. To facilitate an efficient comparison of users based on the above glyphs (**T3**), we developed a novel layout approach based on a triangle mesh to provide a fast and high quality placement of user nodes for supporting interactive data exploration. We design this layout with several considerations: (1) the mesh helps to build a discrete coordinate system, enabling a fast layout with linear complexity and helping with eliminating the node overlaps, (2) triangle mesh, widely used in computer graphics, is able to preserve topologies of any surface or shape. Thus it can be used for capturing the topology of the glyphs' similarity graph or users' interaction graph. Particularly, we place the glyphs on the vertices in a triangle grid and try to maximize the average similarities between neighboring glyphs formally described in the following objective:

$$\mathscr{F} = \frac{1}{|E|} \sum_{(v_i,v_j)\in E} s_{ij} \tag{4}$$

where $E = (v_i, v_j)$ is the collection of neighboring glyphs, i and j, in the triangle mesh and s_{ij} indicates their similarities in the feature space. This objective tends to place users with similar behaviors close to each other based on their feature similarity. It helps to produce patterns such as user clusters, thus facilitating a fast comparison of different groups of users. Fig. 10 shows an example of clusters revealed based on this layout method. The detailed implementation methods and layout evaluation are described in [4].

5.5 Other Contextual Views

Several additional views are also developed, showing different contexts of users' communication behaviors from different angles.

In particular, we developed a message view (Fig. 4.3) illustrating all the raw messages that a focused user posted or responded to during all the communications s/he is involved in (**T5**). In this view, we summarize the high frequency keywords extracted from these messages in a tag cloud, showing the content overview of the messages.

In addition to the feature glyph, we also illustrate the changes of the z-scores of a user's features over time in a temporal heatmap as shown in Fig. 4.5 (**T4**). In this view, each selected user is visualized in an independent heatmap in which rows indicate different features, columns indicate different time points, and each cell indicates the z-score value of the corresponding feature at the corresponding time. The z-score value is visualized by colors ranging from blue to red, showing normalized z-score values ranging from -1 to 1.

Finally, we also employed the design introduced in FluxFlow [45] to illustrate a centric user's impact in terms of message propagation as shown in Fig. 4.6 (**T4**). In this view, we aggregate a user's involving activity threads all together, and illustrate how the corresponding messages are posted or responded to by others overtime. Here, each circle indicates a user, and all the users are packed together in an order determined by the time when each user is involved in these communications. A user who involved in multiple communications is shown in multiple replicas at different places. Consistent with the encoding scheme used in other views, these circles are sized by the users' importance and colored by their anomaly scores or sentiments.

5.6 Interactions

We propose following interactions to efficiently navigate through the data and switch among different information contests.

Query. Using a query box (Fig. 4.4.a), analysts can query to find and load the top ranking suspicious users under a certain topic indicated by a set of keywords.

Filter. A range slider (Fig. 4.4.b) is designed to filter users based their importance, impact, or anomaly scores in a specific range.

Highlight. When hovering the mouse over a user's node in a view, the same user shown in different views is highlighted at the same time, showing the innate connection of the data (**T3**).

Inspection. We support several interactions to support detailed inspection of interesting data items. Particularly, in the global view, the anomalous users can be picked up by mouse clicking or brushing the corresponding nodes. In the inspection view, analysts can select user nodes one by one into contextual views by mouse clicking.

Switch Contexts. Visualizations in TargetVue encode rich contexts which are switchable via interactions. Analysts can switch among different contextual views by clicking the corresponding tabs in the UI. They can also switch among different visual glyphs (i.e., behavior glyph, z-glyph, and relation glyph) and color schemes (i.e, either color by sentiments or anomaly scores) by clicking the corresponding buttons in the toolbar (Fig. 4.4.(c,b)).

Data Labeling. Analyst can label suspicious users and those normal ones via a pop-up menu (**T6**). The suspicious users are highlighted and the normal users are delighted once labeled. All the labeled data will be submitted to the sever and stored for tuning the underlying analysis module.

Zoom and Pan. Both global view and inspection view support zooming and panning for exploring a large set of data items. Analyst can drag the mouse to pan the view and double click or scroll the mouse wheel to zoom.

6 Evaluation

We evaluated the TargetVue system via (1) a social bot detection challenge, (2) a case study using the Enron email data, and (3) in-depth interviews with two domain experts.

6.1 Social Bot Detection Challenge

We applied TargetVue to a social bot detection challenge arranged by our project sponsor. The data of this challenge came from a previous social bot influence challenge held at the end of 2014. The goal of the influence challenge was to design fully automated social bots in Twitter to promote the advantages of vaccination and influence a target network of users who tweet or retweet messages of an anti-vaccine nature. Two teams participated in the influence challenge, which lasted for a month in November 2014. Each team deployed multiple bots. During the influence challenge, the contest organizer collected 100% of the tweets from the deployed social bots and all the users in the target network, as well as user profile and follower/followee information.

Based on the raw data collected, a dataset was created following the influence challenge, which contains about 4 million tweets, 8,000 Twitter accounts selected from the target network containing all social bots and a subset of users, and several snapshots of the follower/followee graph. The snapshots include all (about 46 million) users who followed or were followed by the aforementioned Twitter accounts in the target network. In total there were about 214 million links in a snapshot. During the bot detection challenge, which lasted from mid-February to mid March in 2015, the data from the dataset was replayed in a streaming fashion over time to simulate the live Twitter activities observed during the influence challenge. Several time-sensitive API endpoints were provided for the participants to obtain the data. The snapshots were made available at the beginning of each week.

The bot detection challenge involved six teams (including us) from both industry and academia. The participants were encouraged to try all types of techniques and strategies for detecting the bots as quickly as possible. We used TargetVue system as the primary analysis tool during the challenge. A set of 58 communication features were extracted from the data, guided by the six high-level feature categories discussed in Section 4.2.

Throughout the challenge, all of our four team members worked part-time on it. During that period, we ran the system to automatically collect, pre-process, and analyze the data twice a day as more data came in. The whole process took about 2.5 hours, after which the top 300 anomalous users were visualized (other users could also be acquired by query). We extensively used TargetVue system to detect social bots and created a daily report of suspicious users. The identified bots (i.e. correct guesses) were later used for tuning the analysis model and preparing for the next round of analysis. Many bugs and usability issues were also reported, which were all fixed in time. We successfully identified all the social bots (39) one week before the challenge ended and only had four wrong guesses (false positives). As far as we know, we were the only team employing a visual analytic system, while other teams primarily relied on backend text analytics. In addition, the teams that finished ahead of us spent significantly more person-hours on the challenge.

Explorative Analysis. Fig. 4(1) illustrates the distribution of the top 300 anomalous users from our first anomaly detection results generated by the end of the first week during the contest (we spent a week to collect enough data for the first analysis). At that time, we had no prior knowledge about the data and fully depended on the TargetVue system. The first glance at this global view made two impressions: (1) the anomaly scores computed by the TLOF model were largely consistent with the degree of the outlierness determined by the contour map rendered on top of the MDS projection; (2) the projection results revealed the one-class nature of the problem, i.e., most of the users were densely placed at the center of the view and only a few of them were placed at the marginal area as the outliers (with a high degree of outlinerness). We selected those outlier users with high anomaly scores into the inspection view for a deeper investigation of their behaviors.

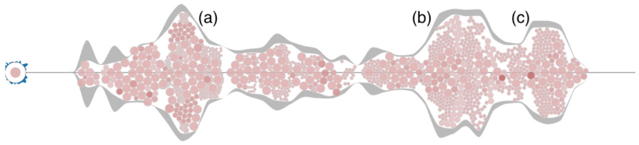


Fig. 7. User (6)'s propagation view showed that most of his responders had high anomalous scores. (a,b,c) illustrate three peaks when (6) influenced the most of the users.

As shown in Fig. 1, in the inspection view, users with similar behaviors were laid out together. Particularly, Fig. 1(a) illustrates the behavior glyphs of the selected users revealing many interesting patterns: (1) *Activeness*. Most of the users behaved actively as their activity threads were densely shown in the corresponding glyphs. In comparison, some of the users such as (4, 5, 6, 18) were quiet with few tweets; (2) *Impacts*. Some of the users such as (16, 19) generated great impact as the tweets posted or retweeted by them spread for a relatively long time shown as the long threads in their behavior glyphs and many other people were also involved in these threads as shown in the Fig. 7. In contrast, some other users such as (2, 7, 10, 17), although very active, had little impact as their posts were rarely retweeted by others; (3) *Sentiments*. All these users had no strong sentiments as most of their tweets were colored in orange (i.e., neutral); (4) *Suspicious behaviors*. Some of the users only posted but rarely retweeted, such as (3, 4, 5, 6, 7, 8, 9, 12, 17). On the contrary, some users only retweeted but rarely posted, such as (1, 11, 15, 16, 17, 19). This indicates two different roles: the message initiator (creator) and the message responder (spreader). A social bot may act in either of these roles. We also found that some users' activities showed a clear periodic pattern such as (6, 11, 15), which were also uncommon behaviors. Selecting them into the feature variance view revealed more patterns. As shown in Fig. 8, user (6)'s periodic posting patterns were more specifically interpreted in the corresponding temporal heatmap. All the z-score values of the features related with the number of urls/hashtags in tweets were periodically changed over time, indicating (6) posted many hash tags and urls, which was not a common behavior of ordinary users.

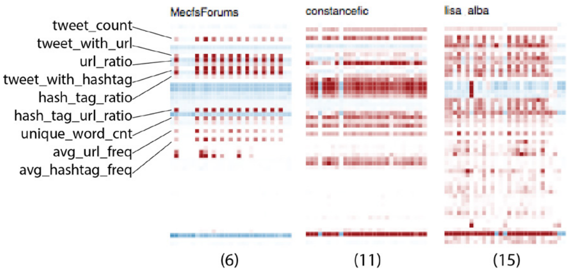


Fig. 8. Temporal heatmaps of three users (6, 11, 15) shown in Fig. 1, illustrating the changing of the z-scores of their features over time.

When switching to the z-glyphs to investigate the users' features, we had more findings. The user (16) who behaved aggressively, as shown in Fig 1(a), had a set of very ordinary feature values. Visually, his z-glyph was the closest to the baseline circle among all the z-glyphs of the selected users. In comparison, some users, such as (3, 4, 5, 8, 9, 14, 18), seemed to be ordinary, but their z-glyphs had very irregular shapes. From the relation glyph view (Fig. 1(c)), we found that most of these users were connected together by following each other. The user (12, 16, 17) had most followers. However, some users such as (3, 4, 5, 9, 10, 14) seem to be isolated.

After inspecting the raw tweets for the aforementioned suspicious users, we found that the users with periodic behaviors were news media accounts that post messages in a regular pace. The users who had lots of followers, such as user (12) or great impact, such as users (16) and (19) were also not social bots as it seemed impossible for a bot to attract so many followers and influence so many people in such a short time period. In addition, those users with both retweeting and posting behaviors such as (1, 2, 18) were also normal users. Finally, we locked our target on users (3, 4, 5, 9, 14), among which users (4, 9, 14) were later verified as social bots.

Tuning the TLOF Model. Based on the lessons and experiences learned from the above process, we tuned the TLOF model in pursuit of better precision. We found the behaviors of the bots that can be grouped into a small number (2 or 3) of clusters, which means inside each cluster we needed to investigate a larger number of behaviors. Therefore, a larger k in Eq. 2 was preferred to cover a larger neighborhood of each behavior in the feature space. Second, we also found that the long-term behaviors of most bots were consistent with themselves; we believe this is because the bots were generated by pre-defined rules and such rules did not vary as frequently as human behaviors. Such observations suggested a small value for the trade-off constant α as well as a shorter time window W in Eq. 1. In addition to tuning these parameters, we also adjusted the weight of each feature based on the features of the bots that had already been found. Based on these strategies, we gradually achieved better and better performance (Fig. 9).

Final Results. The above data exploration and model tuning processes were iteratively performed during the contest. Finally, we successfully found all the social bots with only 4 wrong guesses in total. Fig. 10 illustrates the overview of these bots which are automatically grouped into two clusters, implying two different bot design strategies. These results verified the effectiveness of the features that we selected, i.e., these features successfully separated users with different behaviors. We also found that one group of bots tried to influence others by posting messages. Apparently, this design was not successful at the beginning as most of these bots were suspended by Twitter shortly

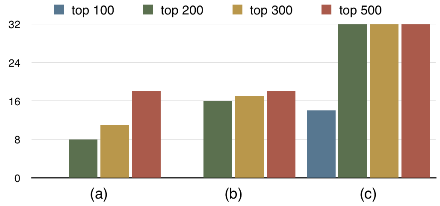


Fig. 9. The total numbers of social bots (heights of the bars) that have been found in top 100 - 500 users ranked by TLOF model based on parameter settings (a) before the contest, (b) after 5 bots were detected, and (c) after about half of the bots (16) were known.

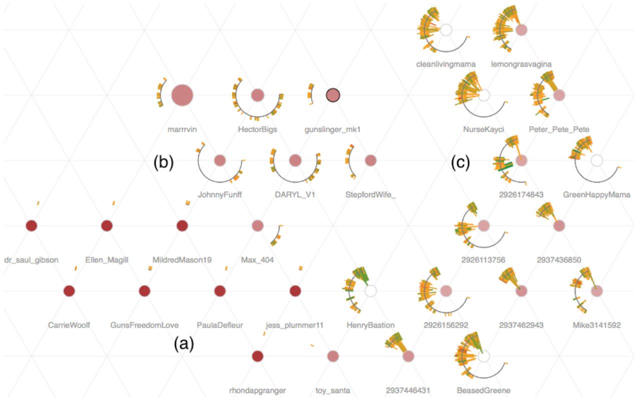


Fig. 10. The behaviors of all the social bots

after they posted messages considered to be spams (Fig. 10(a)). It seemed that the bot designers changed the posting strategy later, which successfully enabled some bots to survive for a longer period time (Fig. 10(b)). In comparison, another group of bots always retweeted messages. This strategy was more successful as most of them had a much longer life circle and were able to retweet a substantial amount of tweets and influence many people.

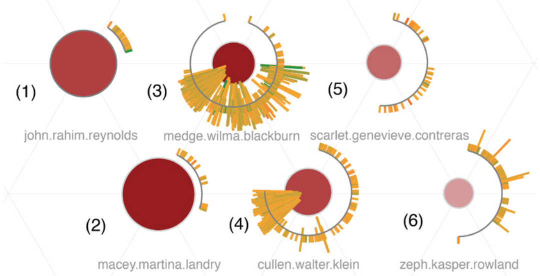


Fig. 11. Some anomalous users in the Enron Email data.

6.2 Interpreting the Email Data

The TargetVue system was also used for analyzing the Enron email data as shown in Fig. 11. In this figure, user circles were sized by the number of unique email accounts the users sent/received emails to/from, and colored by their anomaly scores. When a user replied emails in a thread multiple times, only the first replying time was shown in the glyph as the intersection between the thread segment and the time arc. We can easily interpret the behaviors of the users based on these glyphs. For example, users (1,2) received a large number of emails (with a big circle size) but never replied (no replying threads shown in the glyph); users (3, 4) were involved in a large number of email threads which lasted for a long time. When compared with (3), (4) was often involved near the end of each thread. Users (5, 6) only sent emails but never replied.

6.3 Domain Expert Interview

We performed in-depth interviews with two experts to evaluate the usability of our system. The first expert is one of the organizers of the aforementioned bot detection challenge. The second expert is a research manager whose team is building a security system for a big IT company. Both of them never heard about the TargetVue system before the interview. In each of the interview sessions, we started with a tutorial to explain the purpose and features of TargetVue. We then asked the experts to use TargetVue on their own for detecting anomalous users. After they fully explored the tool's capabilities, we conducted a semi-structured interview, guided by a set of questions. We deliberately asked the experts not to be constrained by the guiding questions, but instead, to use them as a guide and elaborate their thoughts while using the tool. Each of the interviews lasted approximately 1 hour. We recorded the interviews entirely, and took notes of their comments. Both experts were very much impressed by the tool, particularly by the amount of information offered by TargetVue as well as the design itself. They commented that showing different contexts in multiple connected views is "comprehensive", and "very powerful for illustrating different information pieces".

Specifically, the first expert emphasized the novelty of our tool and its importance to the problem of anomaly detection. He said "this is the first time that I am able to directly see users' behaviors". He was very interested in the triangle layout and commented that "putting them [glyphs] in a grid is a smart idea as it facilitates a quick comparison." He also mentioned that "the idea of representing different anomaly measurements in different views is great for comparing different anomaly detection methods." The second expert particularly liked the z-glyph design, the feature variance view, and their layouts. He said: "the z-glyph clearly visualizes how different the users are comparing the whole population, and by combining with the feature variance view, one can easily find new anomalous behaviors that were not noticed before." In addition, he thought using both color and size of the nodes to represent important information (e.g., importance and anomaly score) of each user was intuitive and helped save a lot of time determining which users warrant further investigations. He also thought the behavior glyph design clearly summarized the users' activities: "it is amazing that so much information, time, lifetime, sentiment of a tweet can be packed in such a compact representation." We also asked about the most problematic aspect of our tool. Both experts mentioned that since the visualization incorporates a lot of information about the anomalous users, there will be a bit of a learning curve at the beginning to get familiar with all the views. However they also said, "once you get used to it, the tool is very efficient and comprehensive".

7 Conclusion

In this paper, we propose a novel visual analysis system, TargetVue, for detecting anomalous users via novel visualization designs with multiple coordinated contextual views and a well adopted unsupervised learning model. TargetVue incorporates three new ego-centric glyphs to visually summarize a user's behaviors, which effectively represent the user's communication activities, features, and social interactions. An efficient layout method is proposed to place these glyphs on a triangle grid, which captures similarities among users and facilitates comparisons of behaviors of different users. We demonstrated the power of TargetVue through its application in a social bot detection challenge using Twitter data, a case study based on email records, and an interview with expert users. Currently, tuning the anomaly detection model based on users' feedback is done through a manual procedure. In the future, we would like to design and integrate into the system more advanced methods based on active learning techniques. We also want to conduct a formal user study to further evaluate the usability of our system.

Acknowledgments

This material is based upon work partially supported by the U.S. Defense Advanced Research Projects Agency (DARPA) under Agreement Number W911NF-12-C-0028. The content of the information in this document does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

[1] S. Axelsson. Visualization for intrusion detection. In *Proceedings Computer Security*, pages 309–325, 2003.

[2] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

[3] N. Cao, Y.-R. Lin, F. Du, and D. Wang. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications*, pp(99):1–8, 2015.

[4] N. Cao, Y.-R. Lin, and D. Gotz. Untangle map: Visual analysis of probabilistic multi-label data. *IEEE Transactions on Visualization and Computer Graphics*, pp(99):1–15, 2015.

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

[6] N. Cao, L. Lu, Y.-R. Lin, F. Wang, and Z. Wen. Socialhelix: visual analysis of sentiment divergence in social media. *Journal of Visualization*, 18(2):221–235.

[7] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 10(10-17):30, 2010.

[8] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15, 2009.

[9] E. Corchado and Á. Herrero. Neural visualization of network traffic data for intrusion detection. *Applied Soft Computing*, 11(2):2042–2056, 2011.

[10] R. F. Erbacher, K. L. Walker, and D. A. Frincke. Intrusion and misuse detection in large-scale systems. *IEEE Computer Graphics and Applications*, 22(1):38–47, 2002.

[11] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo. A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security*, pages 77–101. 2002.

[12] B. J. Fry. *Organic information design*. PhD thesis, Massachusetts Institute of Technology, 2000.

[13] B. Gonçalves, N. Perra, and A. Vespignani. Modeling users' activity on twitter networks: Validation of dunbar's number. *PloS one*, 6(8):e22656, 2011.

[14] A. Inselberg and B. Dimsdale. Parallel coordinates. In *Human-Machine Interactive Systems*, pages 199–233. 1991.

[15] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, pages 56–65, 2007.

[16] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[17] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 107–116, 2001.

[18] E. Katz and P. F. Lazarsfeld. *Personal Influence, The part played by people in the flow of mass communications*. Transaction Publishers, 1970.

[19] A. Kind, M. P. Stoecklin, and X. Dimitropoulos. Histogram-based traffic anomaly detection. *IEEE Transactions on Network and Service Management*, 6(2):110–121, 2009.

[20] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[21] J. B. Kruskal and M. Wish. *Multidimensional scaling*, volume 11. Sage, 1978.

[22] S. Kumar, R. Zafarani, and H. Liu. Understanding user migration patterns in social media. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, pages 1204–1209, 2011.

[23] P. Laskov, K. Rieck, C. Schäfer, and K.-R. Müller. Visualization of anomaly detection using prediction sensitivity. 2:197–208, 2005.

[24] W.-J. Li, S. Hershkop, and S. J. Stolfo. Email archive analysis through graphical visualization. In *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pages 128–132, 2004.

[25] J. Lin, E. Keogh, and S. Lonardi. Visualizing and discovering non-trivial patterns in large time series databases. *Information visualization*, 4(2):61–82, 2005.

[26] A. Muñoz and J. Muruzábal. Self-organizing maps for outlier detection. *Neurocomputing*, 18(1):33–60, 1998.

[27] M. Novotny and H. Hauser. Outlier-preserving focus + context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900, 2006.

[28] M. Pennacchiotti and A.-M. Popescu. A machine learning approach to twitter user classification. *Proceedings of the AAAI International Conference on Weblogs and Social Media*, 11:281–288, 2011.

[29] A. Perer, B. Shneiderman, and D. W. Oard. Using rhythms of relationships to understand e-mail archives. *Journal of the American Society for Information Science and Technology*, 57(14):1936–1948, 2006.

[30] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In *Proceedings of the AAAI International Conference on Weblogs and Social Media*, pages 297 – 306, 2011.

[31] N. Simon. *The Participatory Museum*. Museum 2.0, 2010.

[32] Y. Song, Z. Wen, C.-Y. Lin, and R. Davis. One-class conditional random fields for sequential anomaly detection. In *Proceedings of the international conference on Artificial Intelligence*, pages 1685–1691, 2013.

[33] I. Steinwart, D. R. Hush, and C. Scovel. A classification framework for anomaly detection. In *Journal of Machine Learning Research*, pages 211–232, 2005.

[34] S. T. Teoh, K. L. Ma, S. F. Wu, and X. Zhao. Case study: Interactive visualization for internet security. In *Proceedings of the conference on Visualization*, pages 505–508, 2002.

[35] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Proceedings of the IEEE Pacific Visualization Symposium*, pages 41–48. IEEE, 2012.

[36] R. Tinati, L. Carr, W. Hall, and J. Bentwood. Identifying communicator roles in twitter. In *Proceedings of the international conference companion on World Wide Web*, pages 1161–1168, 2012.

[37] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin. Intrusion detection by machine learning: A review. *Expert Systems with Applications*, 36(10):11994–12000, 2009.

[38] B. A. Turlach. *Bandwidth selection in kernel density estimation: A review*. Université catholique de Louvain, 1993.

[39] F. B. Viégas, S. Golder, and J. Donath. Visualizing email content: portraying relationships from conversational histories. In *Proceeding of the ACM conference on Human Factors in computing systems*, pages 979–988, 2006.

[40] F. B. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the ACM conference on Human Factors in Computing Systems*, pages 575–582, 2004.

[41] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the ACM workshop on Online social networks*, pages 37–42, 2009.

[42] R. Xiong and J. Donath. Peoplegarden: creating data portraits for users. In *Proceedings of the ACM symposium on User interface software and technology*, pages 37–44, 1999.

[43] Z. Xu, Y. Zhang, Y. Wu, and Q. Yang. Modeling user posting behavior on social media. In *Proceedings of the international conference on Research and development in information retrieval*, pages 545–554, 2012.

[44] Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su. Understanding retweeting behaviors in social networks. In *Proceedings of the ACM International conference on Information and knowledge management*, pages 1633–1636, 2010.

[45] 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 Transactions on Visualization and Computer Graphics*, 20(12):1773–1782, 2014.