Proceedings of the Seventeenth International AAAI Conference on Web and Social Media (ICWSM 2023)

Motif-Based Exploratory Data Analysis for State-Backed Platform Manipulation on Twitter

Khuzaima Hameed[1], Rob Johnston[2,3], Brent Younce[2], Minh Tang[1], Alyson Wilson[1, 2]

[1]Department of Statistics, North Carolina State University, Raleigh, NC, USA
[2]Laboratory for Analytic Sciences, North Carolina State University, Raleigh, NC, USA
[3]Johnston Analytics, Raleigh, NC, USA
kshameed@ncsu.edu, rob@johnstonanalytics.com, bjyounce@ncsu.edu, minh_tang@ncsu.edu, agwilso2@ncsu.edu

Abstract

State-backed platform manipulation (SBPM) on Twitter has been a prominent public issue since the 2016 US election cycle. Identifying and characterizing users on Twitter as belonging to a state-backed campaign is an important part of mitigating their influence. In this paper, we propose a novel time series feature grounded in social science to characterize dynamic user networks on Twitter. We introduce a classification approach, motif functional data analysis (MFDA), that captures the evolution of motifs in temporal networks, which is a useful feature for analyzing malign influence. We evaluate MFDA on data from known SBPM campaigns on Twitter and representative authentic data and compare performance to other classification methods. To further leverage our dynamic feature, we use the changes in network structure captured by motifs to help uncover real-world events using anomaly detection.

Introduction

Since the 2016 election cycle, state-backed platform manipulation (SBPM) has been a prominent and increasingly important issue in international affairs—specifically, how state actors manipulate social media (Guo and Vosoughi 2020; Zannettou et al. 2019). Social media has been used by governments to promote agendas, spread disinformation, and divert narratives in an effort to influence or manipulate public opinion domestically and abroad.

As a result of recent nation state behavior, the detection and characterization of such actions is essential as part of strategic decision making. Since 2016, the social media platform Twitter has released user, Tweet, image, and video data from various SBPM campaigns originating in Russia, Iran, Venezuela, and other countries. These campaigns have demonstrated platform manipulation with varying levels of coordination. While analysis of individual users is useful, higher levels of coordination suggest this approach is inadequate. Thus our analysis focuses on network-level analysis to detect and characterize malign influence campaigns.

We represent Twitter users as nodes and their replies to each other as edges in a temporal network, which is used to characterize dynamic groups of users engaging in SBPM on social media. We characterize temporal networks using motifs, which are subnetwork patterns that repeatedly occur in a network. Motifs provide a useful way to calculate the "moments" of a network (Maugis, Olhede, and Wolfe 2017). Specifically, motif counts provide numerical summaries of a network, which can be used as features for classification or other analyses.

There are two main contributions of this paper. First, we develop a novel feature that uses motifs to describe the evolution of temporal networks. This feature provides a framework to study SBPM networks on Twitter. Second, we apply this feature to two problems: first to the classification of SBPM networks, and second to uncovering news events through anomaly detection on SBPM networks; finding that SBPM networks appear to have distinct evolution from authentic networks, and that there is a potential relationship between our feature and the occurrence of real-world events.

Methods for network level classification generally use subgraph embedding (Cangea et al. 2018; Hamilton, Ying, and Leskovec 2018). However, the classification of SBPM networks is not well explored, and the optimal features for classifying SBPM networks are not well understood. Our analysis has found the use of static features such as nodes, edges, and density to classify SBPM networks to be inadequate. To motivate our approach, we leverage the idea that the evolution of networks differs when they are used to spread misinformation (Juul and Ugander 2021), which is a common tactic of SBPM campaigns (Friggeri et al. 2014; Vosoughi, Roy, and Aral 2018). This difference can manifest in the structure of the network summarized by network motifs. We specifically consider motifs that characterize central network structure, which we hypothesize is a useful perspective for SBPM networks. To study this, we summarize temporal motif counts with a matrix, and then perform classification to analyze the difference between SBPM and non-SBPM data in terms of central network structure. Moreover, temporal motif counts provide us with a real-time view of dynamic networks. This leads us to perform time series analysis, namely anomaly detection to detect real-world events.

Looking more broadly, social media networks are also studied in anthropology, economics, psychology, and political science. These fields have explored the various mechanisms involved in communicative behavior online. From anthropology, the concepts of identity (Mach 1993), power (Wolf 1999), and reciprocity (Mauss 2002) are used to ex-

plain the social behavior of individuals, specifically the notion that there is intangible value attributed with acts of offering service to others. This is used to help explain the motivation of behavior of communities online (Skageby 2010). For example, Chakrabarti and Berthon (2012) describe the goal of social media influencers to provoke desired responses from individuals by eliciting various emotions. An example from their paper was user backlash on Facebook from Nestlé's response to allegations of environmental harm through their use of palm oil. The backlash manifests as a change in engagement toward Nestlé, which highlights the effect of influencers on dynamic social networks. This supports our use of dynamic network features in understanding the effect of influence campaigns. This in turn can help attribute tangible phenomena to the intangible value of reciprocity online.

Behavioral economics often studies irrational behavior in decision making; for example, individuals act on their biases when exchanging social currency (Thaler 2016), such as the tendency of individuals to organize in like-minded groups (Knobloch-Westerwick, Mothes, and Polavin 2020). This supports the tactic of SBPM users arranging themselves in distinct groups to target corresponding groups of users (Linvill and Warren 2020). Linvill and Warren (2020) also reveal varying behavior across groups, in part demonstrated by their behavior over time. This suggests dynamic network features support the differentiation of groups of SBPM users online.

There are also political motivations for the use of social media. For example, Nye (2002) discusses the notion of soft power and how it helps to explain the motivation of state-sponsored influencers associating with or impersonating reliable sources of information, while Moghaddam (2005) discusses the impact of radicalization and how it gradually indoctrinated individuals to commit various offenses. In the online space, Gill et al. (2017) found groups of extreme-right-wing individuals were over four times more likely to use the web to prepare for attacks than Jihadist-inspired individuals. This highlights the increasing utilization of the web by extremists, and the opportunity afforded to state-backed actors to target these individuals. Gill et al. (2017) also suggest an increased focus of intervention on tools for radicalization, in contrast to the portrayal of radicalization itself. The increasing role of the web in the gradual radicalization of individuals can benefit from studying temporal SPBM networks that often engage in radical messaging.

Related Work

Analysis of SBPM Data

There are a wide variety of existing analyses of SBPM data from Twitter. Most analyses focus on user-level analysis, i.e., the analysis of individual users or their content. Such analyses use user-level information for classification (Badawy, Lerman, and Ferrara 2019; Ferrara 2017), to classify Russian trolls using text (Ghanem, Buscaldi, and Rosso 2019), to construct behavioral attributes for Russian troll classification (Luceri, Giordano, and Ferrara 2020; Alhazbi 2020), to study the changes in platform usage behavior (Badawy, Lerman, and Ferrara 2019; Bail et al. 2020; Linvill and Warren 2020), or to combine different features for classification (Im et al. 2020). Other methods focus on features such as network structure over time (Badawy et al. 2019), the comparative numerical attributes of a Russian troll network (Stewart, Arif, and Starbird 2018), or influence detection through a stochastic process model (Zannettou et al. 2019).

The analyses mentioned thus far use a limited amount of SBPM data, primarily from Russia and Iran. Since there have been a variety of countries involved in SBPM in recent years (e.g., Venezuela and Bangladesh), we believe that there is a need to expand the data used to characterize SBPM behavior. Indeed, text, behavior, and networks vary across campaigns from different countries (Bradshaw and Howard 2017; Woolley and Howard 2017, 2018; Beskow and Carley 2020; Vargas, Emami, and Traynor 2020; Alizadeh et al. 2020). Thus, there is a concern that the classification approaches mentioned so far would have difficulty in maintaining their performance with newer data or across different campaigns.

Badawy, Lerman, and Ferrara (2019) address this issue directly, discussing how their limited training data is problematic for classification on a wider set of users. One shortcoming they identify is that liberal users are not studied as thoroughly because the Russian campaigns did not target them. Such examples illuminate the difficulty that comes with understanding SBPM campaigns. In addition, there is a limitation in identifying users in comparison to the identification of coordinated groups of users engaging in a malign campaign. Indeed, there are instances where individual-level classifiers are inadequate in discovering platform abuse (Grimme, Assenmacher, and Adam 2018).

Badawy, Lerman, and Ferrara (2019) also study the centrality (or relative importance) of users in SBPM networks over time, which has similarity to our work in that we study the structural attributes of SBPM networks as a whole. If SBPM users have unique temporal network attributes, it supports our hypothesis that SBPM networks have distinct dynamics with respect to central actors compared to authentic networks.

For exploratory network-level analysis, Linvill and Warren (2020) considered three different edge types from Internet Research Agency Twitter data, while studying six different manually labeled categories of users based on their political affiliation and behavior. They show that users tend to communicate more strongly with others from the same category. However, they provide no classification analysis of network data.

Classification of SBPM Campaigns There is little work on the classification of SBPM data at the network level. Vargas, Emami, and Traynor (2020) analyzed 10 strategic information operations (SIO) campaigns (which we call SBPM) from Twitter. Instead of focusing on classifying SIO campaigns, they focused on classifying SIO-like activity, a more narrow target than ours. They calculate network statistics from six different edge types, then concatenate them together. They split the data by two different periods to predict

future SIO-like activity. They also perform classification of campaigns as a whole. However, they have difficulty with the Twitter data having varied behavior within and across campaigns. In addition, their large feature set may make their model prone to over-fitting.

Alizadeh et al. (2020) look at Tweet-based features to identify campaigns using Russian, Chinese, and Venezuelan SBPM data, and they use period-splitting similarly to Vargas, Emami, and Traynor (2020). Martino et al. (2020) discuss detecting coordinated, inauthentic activity using binary classification, and they highlight the issue of using potentially outdated data from bots or campaigns (e.g., the work of Cresci et al. (2017) with bots). Their work also notes that recent approaches tend to either explore different classification targets or simply avoid classification in favor of unsupervised learning. Pacheco et al. (2021) studied classifying coordinated behavior from a Hong Kong protest dataset, as opposed to identifying SBPM directly.

Vargas, Emami, and Traynor (2020) is most similar to our approach, in terms of the breadth of datasets used and performing network classification. However, their method only studies SBPM networks from a static perspective. By using time series features, we can study SBPM campaigns from a dynamic perspective, and in our case perform anomaly detection.

Anomaly Detection in Social Media

There is increasing interest in linking anomalous events in social media data to real-world events, and anomaly detection is a common method used for this task. Anomaly detection can be viewed as the problem of finding observations that are outliers with respect to the generative parameters of a time series. This may include parameters associated with the mean or covariance function of a time series.

In addition to the algorithms used for anomaly detection, the features constituting the time series data are also important for this problem. Social media attributes that have been used include friend groups (Fire, Katz, and Elovici 2012), text topics (Lauschke and Ntoutsi 2012), replies and retweets (Takahashi, Tomioka, and Yamanishi 2014), keywords, Tweet counts, and mentions (Takahashi, Tomioka, and Yamanishi 2014; Hendrickson et al. 2015), and revenue per engagement (Zhang et al. 2015).

More generally, various attributes of dynamic networks have been used to detect anomalies. Attributes include anomalous groups within the network (Miller, Arcolano, and Bliss 2013; Mongiovi et al. 2013; Yu, He, and Liu 2015), network scan statistics (Cheng and Dickinson 2013; Neil et al. 2013), magnitude of network similarity measures (Bunke et al. 2007; Rossi et al. 2013), simple network statistics (Kendrick, Musial, and Gabrys 2018), or a probabilistic model (Heard et al. 2010; Bhamidi, Jin, and Nobel 2018). Descriptive surveys of anomaly detection are provided in (Savage et al. 2014; Akoglu, Tong, and Koutra 2015; Ranshous et al. 2015; Yu et al. 2016).

Motifs have a distinct role in detecting anomalous network activity. Namely, the 3-path and 4-star motifs that we analyze are related to centrality characteristics within the network. If counts of these motifs increase, then it suggests increased engagement to and from central actors. Thus if an event occurs in the real world, it can coincide with an increase in 3-path and 4-star motif counts.

Data

Twitter

For the classification problem, our data fall into two classes, which we label *inauthentic* or SBPM and *authentic*. For inauthentic data, since 2016 Twitter has collected and periodically released content from accounts they have associated with SBPM[1]. According to Twitter, accounts tagged this way were involved in platform manipulation that they can reliably attribute to a government or state-backed actor, which violates their terms of service. Examples of platform manipulation under their policy include inauthentic promotion of users or content, influencing conversations through the use of coordinated fake or real accounts, or coordination that violates Twitter rules[2]. From the data released by Twitter, we use a purposive sample of nine datasets attributed to the Internet Research Agency, Russia, Iran, Venezuela, and Bangladesh since 2018. These data include Tweets, users, photos, and other media, along with their metadata. In total, 150,046 users are contained in our SBPM data.

We use three sources for the authentic Twitter data. First, we pull a random sample of Twitter. This data allows us to compare SBPM data with a representative sample of Twitter conversations. However, since authentic campaigns often exhibit behaviors similar to SBPM, such as high levels of coordination, we over-sample data around highly advertised events. Using BrandWatch's (formerly Crimson Hexagon) proprietary algorithm, we collected data around the Night of too Many Stars television event and the 2020 NBA All-Star game. In total, 107,564 users are contained in our authentic data. We note the possibility that there are SBPM users or SBPM activity contained within the authentic data. However, we expect this data to be dominated by authentic activity.

We use a developmental tool, Social Sifter (Johnston, Watts, and Younce 2020), that analyzes coordinated SBPM campaigns using a variety of machine learning models. The tool is used to query Twitter to retrieve authentic data that match the topics of discussion in the SBPM data. We perform 40 queries from this source to supplement the authentic data from other sources. The querying process is described in more detail in the following section.

Organization by Topic We aim to control differences between our data that might impact the development of our classification algorithm. One relevant difference arises from differences in the topics under discussion. To mitigate potential bias from differences in network structure between dissimilar topics, we pulled authentic data with similar topics to those contained in the SBPM data.

[1] https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter

[2] https://help.twitter.com/en/rules-and-policies/platform-manipulation

To construct a topic network, we first identify the top keywords from each SBPM dataset, excluding stop words. Each keyword corresponds to a topic. Next, the top topics are used to query the SBPM and authentic data. First, we query the Tweets for the topic by matching text. Then, the 100 most recent Tweets are pulled from the users contained in the initial query. This two-step process produces the topic network. If a topic is not present in authentic data from the random sample or BrandWatch, we use Social Sifter to query the topic on Twitter. Note there may be multiple networks belonging to a topic.

The training data contains the following topics from both SBPM and authentic sources, translated here from their original language: Khaleda Zia, @banglanews24com, Colombia, Israel, Obama, Pakistan, police, #releasethememo, @rt_russian, Trump, USA, Ukraine, Pakistan, Rohingya, election, Russia, @nicolasmaduro, @forocandanga, Chávez, Clinton, France, Holland, Iran, ISIS, Korea, and Imran Khan. In total, there are 27 authentic and 25 SBPM training networks.

The testing data contains the following topics for authentic networks, again translated from their original language: Twitch, QAnon, fortnite_primal, #EGE2021 (standardized exam in Russia), Atletico_MG (soccer team), Blackrock, data_science, NCSU, Netanyahu, recipes, a port accident in Taiwan, #serverless, #CoronaFromUSA, and @ArmsWatch. For SBPM networks we use the following topics: MAGA, Islamic, Trump, Islam, @ArmsWatch, #impeachmentbackfire, #CoronaOutbreakInIranIsUnderControl, and #CoronaFromUSA. In total, there are 12 authentic networks and 13 SBPM networks in the testing data.

We construct the networks used in Section as follows. We first extract the timestamp and text of Tweets, as well as the replied to user(s) if they exist. This yields a list of edges composed of the author of the Tweet and the users replying to the Tweet. The edge list is used to construct the temporal network, with duplicate and self edges dropped. In our early analysis, we found that retweet network data is very limited in the SBPM data. As such, we do not analyze retweet networks in this paper.

Pre-processing

Once we have topic networks, additional processing is performed to prepare the data for classification and anomaly detection. The format of the data provided by Twitter and the Twitter API is tabular. We first query the ID, username, timestamp, and text of each Tweet. Any reply users are parsed from the Tweet text. Any Tweets that contain a missing ID, username, timestamp, or text are excluded from our analysis. We remove 16 Tweets from this step.

Normalizing Motif Counts Since there are a variety of sizes and densities of reply networks engaged in state-sponsored influence, we normalize motif counts to reduce their dependence on network size and density using a quantity that depends on both of these factors. Let G be a network with n nodes and k edges. Let M be a motif with m nodes and ℓ edges, and let C_m denote the complete network on m nodes.

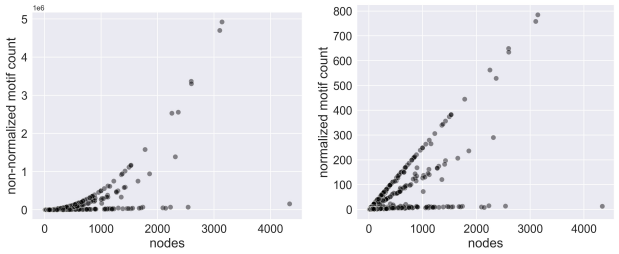

Figure 1: Scatterplots comparing node count and motif count for training reply networks. On the left, no normalization is applied to motif counts. On the right, density-based normalization is applied.

Define the density-based normalization as

$$\text{DBN}(G, M) = \binom{n}{m} |\text{iso}(M, C_m)|\, p^{\ell}, \tag{1}$$

where $p = k/\binom{n}{2}$ is the density of G, and $|\text{iso}(M, C_m)|$ is defined in Equation 3. This is the number of motifs that we would observe under the network model of edges appearing uniformly at random, given the density of the network.

Figure 1 plots motif counts against network size in nodes. This shows the effect of normalization on motif count with respect to network size. Normalized motif counts grow linearly with the number of nodes, allowing for easier comparison of networks of varying sizes and densities. Note there still exists a dependence between network size and motif count after normalization, because networks can deviate from the null model in (1). We also note the fork pattern in both the normalized and non-normalized data, which is likely due to the different network densities in the training data.

Motif Functional Data Analysis Classifier For classification analysis, we have an additional pre-processing step. For a set of Tweets belonging to a particular topic and dataset, we split the data by varying time period lengths to augment the training data and leverage an ensemble method. To allow for a sufficient number of data points to apply the spline learning algorithm in Section , we drop periods with fewer than 14 Tweets overall. Due to the computational cost of counting motifs at multiple time points over many networks, we truncate periods to the first 1000 Tweets overall.

Anomaly Detection The CAPA anomaly detection algorithm requires additional processing of our data. One requirement is that observations are spaced evenly in time. To impute missing values, we fill days with no Tweets using linear interpolation. Another assumption is that the data are stationary in mean and variance. To remove the mean trend, we use a Savitzky-Golay filter (Savitzky and Golay 1964). To remove the variance trend, we perform a Box-Cox transformation, with parameter fitting done for each topic dataset (Box and Cox 1964) using the Python package *scipy* (Virtanen et al. 2020).

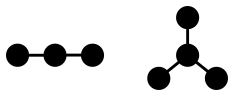

Figure 2: Visualizations of the two network motifs we consider, the 3-path, and 4-star. High occurrences of these motifs indicate a high amount of interaction around central actors in a network.

	3-path		4-star	
	Authentic	SBPM	Authentic	SBPM
count	895	1718	895	1718
mean	214	1640	2564	44823
std	674	2332	30763	100644
min	0	1	0	0
25%	11	214	4	680
50%	98	695	158	3776
75%	276	1840	632	24790
max	12858	14878	659398	848046

Table 1: Summary statistics of motif counts for weekly period-based reply network data.

Figure 3: A thread of three users from Twitter demonstrating how a motif is formed. Karen's initial reply to Daniel produces an edge. Daniel further replies to Karen, but this does not add an edge since one already exists between the two. Spencer's reply to Karen produces another edge, so a 3-path motif is formed.

Methods

Network Motifs

A network motif is a sub-network or pattern. Figure 2 shows the two types of motifs we consider in our analysis, which are the 3-path and 4-star. Motifs represent micro-network interactions between users on social media, where nodes represent users, and edges represent some kind of interaction, like a reply or retweet. Figure 3 gives an example of a Twitter thread containing a 3-path motif from Figure 2. The user Karen authors a Tweet, which prompts replies from two other users. Treating their replies as edges yields the 3-path motif. In general, motifs are not constrained to come from a single thread.

In our early analysis, we explored using all 3- and 4-node motifs. We found that the 3-path and 4-star motifs had the largest presence in our data. In addition, the other 3- and 4-node motifs had a considerably low presence (several of them are non-existent in many of the networks in our data). To allow our method to generalize to a wider variety of network sizes and densities, we use only the 3-path and 4-star motifs. Also, we do not consider motifs with five or more nodes since our training data contains relatively few in number compared to 3- and 4-node motifs. Since we split the data by period for some of our analyses, we provide summary statistics on the count of 3-path and 4-star motifs observed each week. Table 1 provides summary statistics for these data grouped by label. The scale of motif counts appears to differ between SBPM and authentic networks, which supports the use of normalization in Section .

We now formally introduce motif counts. First, let $\mathcal{G}$ be the set of simple *undirected* networks, where a simple network $G \in \mathcal{G}$ is defined as a tuple (V_G, E_G) with vertex set V_G and edge set E_G. We will assume that, for each G, the edges of G have time stamps $\{x_t \colon t \in E_G\}$. For ease of exposition, we will assume that these $\{x_t\}$ are ordered, i.e., $x_1 \leq x_2 \leq \cdots \leq x_{|E_G|}$. We formally define a motif as a simple network M, with number of vertices small compared to the vertices in G. Multiple copies of M may appear in the network G. More specifically, we count the number of edge-preserving isomorphisms from M to G. Denote $\hom(M, G)$ as the set of edge-preserving homomorphisms from M to G, i.e.

$$\hom(M, G) = \{f : (x, y) \in E_M \Rightarrow (f(x), f(y)) \in E_G\}. \tag{2}$$

Define the equivalence class

$$\text{iso}(M, G) = \{\hom(M, G) : f \sim g \Leftrightarrow f(M) = g(M)\}. \tag{3}$$

The motif count of M is defined as $|\text{iso}(M, G)|$. Using the equivalence class ensures that we do not count a motif multiple times if over the same set of nodes.

To count motifs in a network in $\mathcal{G}$, we take as input the list of degrees of each node. For each node i and its degree d_i in the list, we compute the quantity $\binom{d_i}{2}$. This is the number of 3-path motifs centered at the given node. Summing over the motif counts of each node, we get the number of 3-path motifs for the network. Similarly for 4-star motifs, the quantity $\binom{d_i}{3}$ is computed at each node.

To analyze the complexity of our method, consider $G \in \mathcal{G}$. Provided an edge list, we can read it in to compute the degree of node, which requires $|E_G|$ iterations. We then compute the quantities $\binom{d_i}{2}$ and $\binom{d_i}{3}$ for each node i and sum, requiring an additional $|V_G|$ iterations. Thus the total complexity is $O(|E_G| + |V_G|)$.

For our analyses, we are given data in the form of the networks $G_1 \ldots, G_N \in \mathcal{G}$ and associated labels $Y_1, \ldots, Y_N \in \{0, 1\}$, where a network is labeled "0" if it is authentic and "1" if it is inauthentic. Define $\{x_t^{(i)}\} \subset \{1, 2, \ldots\}$ as the set

of ordered timestamps for G_i. Using $\{x_t^{(i)}\}$, we further split each network into time periods of equal length. Let the set of periods $\{P_1^{(i)}, \dots P_{\ell_i}^{(i)}\}$ be a contiguous partition of $\{x_t^{(i)}\}$, where ℓ_i is the number of periods in $G_i, i \in \{1, \dots, N\}$. For network i and $j \in \{1, \dots, \ell_i\}$, denote G_{ij} as the subnetwork induced by $\{x_t^{(i)} \in E : t \in P_j^{(i)}\}$.

For a motif M and each network G_{ij}, we will construct a time series. Let $k \in P_j^{(i)}$, and denote G_{ijk} as the subnetwork induced by $\{x_t^{(i)} : t \in P_j^{(i)}, t \leq k\}$. Define the time series sequence $\text{TS}(M, G_{ij})$ of G_{ij} as

$$\text{TS}(M, G_{ij}) = (|\text{iso}(M, G_{ijk})|)_{k \in P_j^{(i)}}.$$

For convenience, we will denote $\text{TS}(M, G_{ij})_m = (|\text{iso}(M, G_{ijk})|)_m$, where $m \in P_j^{(i)}$. This represents the number of motifs of type M in G_{ij} at a given timestamp in $P_j^{(i)}$. Each time series $\text{TS}(M, G_{ij})$ will be treated as a single data point for our analysis. In total, there are $N_{\text{total}} = \Sigma_{i=1}^{N} \ell_i$ data points. Labels for subnetworks are inherited accordingly, where, for each $i = 1, 2, \dots, N$ we assign $Y_{ij} = Y_i$ for all $j \in \{1, \dots, \ell_i\}$.

Motif Functional Data Analysis Classifier

We describe our classification method based on temporal motif counts. We initially considered gl2vec (Tu et al. 2019), a recent off the shelf motif method. However, the motifs they used had an extremely low presence in our network data, which led to poor classification performance. We do not explore this method further in this paper.

Our classification approach first treats each motif count time series as functional data. Using techniques from functional data analysis, it then represents the functional data in Euclidean space. We call this process Motif Functional Data Analysis (MFDA). We use a generative additive model (Hastie and Tibshirani 1986), implemented in the Python package *pygam* (Serven et al. 2018). Suppose we have the data $\text{TS}(M, G_{11}), \dots, \text{TS}(M, G_{N\ell_N})$. We model each time series $\text{TS}(M, G_{ij})_k$ using the linear model

$$\mathbb{E}[\text{TS}(M, G_{ij})_k] = \beta_1^{(ij)} s_1(t_k) + \cdots + \beta_d^{(ij)} s_d(t_k),$$

where $s_1, \dots, s_d$ are B-spline basis functions. Note that MFDA does not attempt to precisely model $\text{TS}(M, G_{ij})$, but rather perform dimension reduction of the time series (Wang, Chiou, and Müller 2016); this explains the use of real-valued basis functions even when $\text{TS}(M, G_{ij})$ is integer valued.

The coefficient vector $\beta^{(ij)}$ is estimated using the following procedure. Let

$$S^{(ij)} = \begin{bmatrix} s_1(t_1) & \cdots & s_d(t_1) \\ \vdots & \ddots & \vdots \\ s_1(t_k) & \cdots & s_d(t_k) \end{bmatrix}$$

be the data matrix. Let $\widehat{\beta}^{(ij)}$ be the minimizer to the objective function for the B-splines regression given by

$$\|\text{TS}(M, G_{ij}) - S^\top \beta^{(ij)}\|^2 + \lambda (\beta^{(ij)})^\top P \beta^{(ij)},$$

where P is the P-spline penalty matrix (Marx and Eilers 1999) commonly used for spline regression, such that

$$(\beta^{(ij)})^\top P \beta^{(ij)} = \sum_{m=2}^{d} (\beta_m^{(ij)} - \beta_{m-1}^{(ij)})^2,$$

and where λ is a smoothing parameter. We choose λ and d by cross validation.

Define $X = [X_{11} \cdots X_{NN_N}]^\top$ as the $N_{\text{total}} \times d$ covariate matrix, where $X_{ij} = \widehat{\beta}^{(ij)}$ is represented as a row vector. Using this matrix with the corresponding label vector, we can fit any off-the-shelf classifier. We use a random forest. Given the learned classification model $\widehat{f}$, then we can get a classification score for network i with the simple ensemble

$$\frac{1}{N_i} \sum_j \widehat{f}\,(X_{ij}),$$

which is what we use in our evaluations. Do note that as a result of this ensemble, the choice of classification threshold is not obvious. This motivates the use of a classification metric in our evaluations that incorporates all thresholds.

To produce input data for the classifier, we first, for each topic dataset, split the Tweets by period using timestamps. For each period, we sort Tweets chronologically. For each edge added to the network, we count the number of motifs up until the time of the last edge. This produces a time series of motif counts with time steps at each Tweet. Normalization is then applied as described in Section . With the normalized motif time series, we fit the spline model from this section. The resulting coefficient vector and label are input into the classifier.

Figure 4 visualizes part of the process, first plotting the motif time series for a sample of weekly periods from one of the topic networks. Motif counts are normalized using the density-based normalization of Section . Time is scaled to be between 0 and 1, where 0 corresponds to the first Tweet in the period and 1 corresponds to the last Tweet in the period. In the adjacent scatter plot, we apply the remaining steps of the spline learning algorithm to the data to get spline coefficient vectors. We then plot the first two principal components.

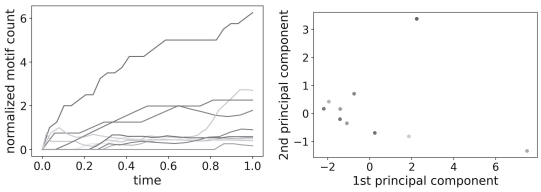

Figure 4: Example of data processing used in Section for Venezuelan Trump topic network. The plot on the left shows density-based normalized motif time series computed from the Venezuelan network discussing the topic of Trump. On the right is a PCA plot of the spline coefficient vectors of the same data.

Anomaly Detection

For anomaly detection, we are given the same networks $G_1, \ldots, G_N$, but with a key difference from the classification problem. We do not split the data as we are analyzing each dataset as a whole, which corresponds to the time series data $\text{TS}(M, G_1), \ldots, \text{TS}(M, G_N)$. The CAPA algorithm (Fisch, Eckley, and Fearnhead 2018) implemented with the *anomaly* package in R (Fisch et al. 2020) first models contiguous segments of $\text{TS}(M, G_i)$ as normal distributions, where each segment has constant mean and variance. The number of segments are chosen through optimization. Roughly speaking, we are looking for a collection of time points $\{t_k\}$ in $\text{TS}(M, G_i)$ such that the motif counts $\text{TS}(M, G_i)_{t_k}$ are outliers with respect to the variance of the data distribution. The data and method we use to evaluate anomaly detection is described in Section .

Evaluations

MFDA

There are five sets of features we compare in classification performance, each of which is computed for each periodic network. The first and second feature sets are composed of the normalized and non-normalized MFDA matrices, as defined in Section . Furthermore, for both the normalized and non-normalized cases, three feature combinations are considered: the MFDA matrix computed from the 3-path motif, the matrix from the 4-path motif, and the matrices from both motifs concatenated horizontally. The third feature set is simply the normalized counts of the 3-path and 4-star motifs contained in the periodic network as a whole. We call this feature set static motif since it does not utilize the temporal information in the network like in MFDA. The fourth feature set is composed of five simple network features: the number of nodes, number of edges, number of components, density, and transitivity of the network.

We also consider an outside approach, with a feature set that is adapted from the coordination network features outlined in Table 2 from the work of Vargas, Emami, and Traynor (2020). They use networks different from reply networks, such as retweet networks. However, due to the lack of retweets in our data, a plain implementation of their features would be sub-optimal. Since we only consider reply networks in this paper, we adapt their method by computing their seven coordination network features on our network data. We also note that they also perform period splitting for the periods of one day and one week, but for a more direct comparison to our method, we use the same periods that are used in the previously mentioned feature sets.

There are nine different feature combinations in total. The network data are split using three different periods: two weeks, one week, and three days. Each type of feature is computed for each time period, which yields 27 total feature matrices. Note that once data are split by, for example, one-week periods, the order of those weeks is not considered when training the model. The same applies to the other period lengths.

The classifier we use is a random forest model with the Gini criterion, with the parameters 100 trees, three minimum samples per leaf, and the remaining parameters having the default values from *scikit-learn* (Pedregosa et al. 2011). Section describes how the training and test data are created. For each feature choice, e.g., normalized MFDA with the 3-path motif, we train three models, one for each period length. Feature matrices are similarly computed for the test networks. For each test network, we produce a vector of outputs from the random forest model. Finally, the vector elements are averaged to produce a classification score for the test network. As alluded to earlier, the classification evaluation metric we use is the area under the receiver operating characteristic curve (AUC), which allows us to consider all thresholds of the classification score.

Table 2 shows the results of the analysis. Each cell contains the AUC computed using the 24 classification scores of the test networks. Overall, no feature combination performs uniformly the best across all period lengths. Static motif counts perform well when the period length is three days, suggesting that its less complex feature set suffices for smaller networks. The coordination network features perform well across all period lengths, in particular for one-week periods. For MFDA, there is a clear benefit from normalization when comparing its AUC scores to its non-normalized counterpart overall. Also, using both motif types in normalized MFDA leads to slightly more consistent performance in comparison to using a single motif type. The strength of considering multiple motif types is also reflected in the AUC scores of static motif counts. Simple network features appear to be inadequate for SBPM classification.

Figure 5 provides a closer look into the classification behavior from three leading feature combinations from Table 2: normalized MFDA using both motif types, static motif, and coordination network. The figure contains box plots of ensemble classification scores grouped by each true label and feature combination, and this is repeated for the three period lengths. Notice that the models trained on static motif and normalized MFDA features tend to score SBPM networks higher than the models trained on coordination network features. On the other hand, models trained on coordination network features tend to score authentic networks relatively low. This suggests that models trained on coordination network features will have a higher false negative rate, whereas models trained on static motifs and normalized MFDA features will have a higher false positive rate. We view the latter quality as preferable for our context, as more SBPM networks would be identified for further analysis.

To study this quality further, for each of the three feature combinations and the three period lengths, we compute the maximum F_2 score across all possible thresholds. The F_2 score weights recall twice as much as precision compared to the usual F_1 score, which is more suitable for our context. The F_2 scores are in Table 3, which shows that given their optimal thresholds, the performance of the three feature combinations becomes more comparable. In summary, MFDA shows utility for SBPM classification, and validates the presence of temporal differences between SBPM and non-SBPM networks.

We conclude the classification analysis by briefly study-

Feature Set	AUC Score		
	3 days	1 week	2 weeks
Non-MFDA			
static motif	0.92	0.85	0.81
simple	0.40	0.52	0.60
coordination network	0.89	0.90	0.87
Normalized MFDA			
3-path	0.86	0.78	0.89
4-star	0.80	0.67	0.90
both	0.90	0.81	0.86
Non-normalized MFDA			
3-path	0.83	0.55	0.77
4-star	0.69	0.56	0.69
both	0.84	0.72	0.89

Table 2: AUC scores from random forest classifier trained on various feature sets and period lengths.

Feature Set	F_2 Score		
	3 days	1 week	2 weeks
static motif	0.89	0.85	0.87
coordination network	0.92	0.89	0.90
normalized MFDA, both	0.95	0.87	0.86

Table 3: Maximum F_2 scores for three leading feature combinations from Table 2 and for each period length.

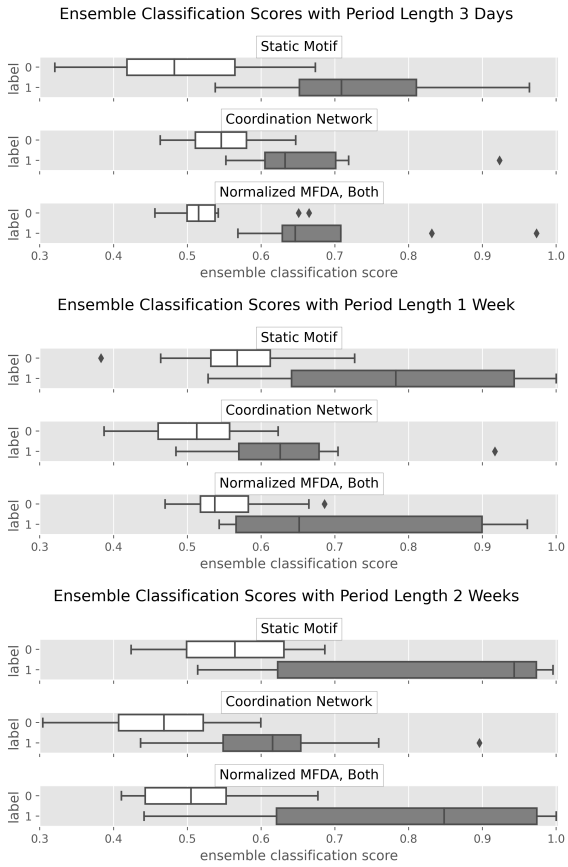

Figure 5: Box plots of ensemble classification scores grouped by the true label and three leading feature combinations from Table 2: normalized MFDA using both motif types, static motif, and coordination network. This plot is repeated for each of the three period lengths.

ing the time series component of MFDA, and specifically by comparing features of motif time series between classes. This provides us with a better understanding of what time series features are responsible for the performance of MFDA. We first compute the normalized 3-path time series for each topic network in the training data, as defined in Section . The time series features we compute for each of the 52 training networks are amplitude, maximum, minimum, mean, median, max slope, standard deviation, percent of data beyond one standard deviation, maximum distance to the median, number of local maxima, the Stetson K statistic, and the sample skewness. Using the features from the 52 networks, we perform a (two-sided) Mann–Whitney U test comparing the ranking of SBPM and non-SBPM time series features. Features corresponding to significant tests are listed in Table 4. These features are the percent of time points that are beyond one standard deviation of the data, the Stetson K statistic (a robust measure of kurtosis), and sample skewness (a measure of the symmetry of the time series values). The median values of these three features are higher for SBPM time series.

Anomaly Detection

A common way of evaluating anomaly detection is to compare detected dates of anomalies to dates of events related to time series data, where our time series data are calculated from the training topic networks described in Section . Note we identify dates of events before calculating anomalies. To do this, we first identify the time periods spanned by each of the topic networks. We then search each topic for major news events within its respective time period, recording the dates of each event. Of all the days across topic networks, 11% of them contain major events we recorded. This can be thought of as the base rate of event detection if the dates of anomalies are chosen at random.

With news events recorded, we then calculate the dates of anomalies in the training topic networks using the CAPA algorithm. To compare the performance of our approach for anomaly detection against an existing approach, two types of time series are calculated for each topic network. The first type is normalized motif time series, constructed as described in Sections and . The other type is constructed as the cumulative counts of Tweets per day in the topic network. Note the processing described in Section is applied to both types of times series. We compare these approaches to the base rate of detection. Figure 6 shows examples of anomaly detection on the two types of time series data, with the strongest anomaly marked in each time series. Notice that for the Tweet count time series plots, the dates of the

feature	statistic	p-value	$m_{\text{SBPM}} - m_{\text{NSBPM}}$
pct. beyond 1 std.	154	0.0061	0.14
Stetson's K	145	0.0042	0.16
skew	87	<0.0001	1.45

Table 4: Two-sided Mann–Whitney *U* test of time series features between SBPM and non-SBPM time series data. The difference in medians between SBPM (m_{SBPM}) and non-SBPM (m_{NSBPM}) is provided for reference.

strongest anomalies detected are near the end of their respective periods. This pattern persists for other topic networks, and it is partly due to the Twitter query API prioritizing recent Tweets. By doing so, each network grows quickly and thus evolves more rapidly toward the end of the period. This again supports the use of normalization, which counters the effect of network size. For a more complete picture in this evaluation, we consider the top ten anomalies from each time series type, ranked by their *p*-values calculated from the CAPA algorithm.

We now discuss how the dates of recorded events compare to the dates of detected anomalies. With motif time series, three of the top 10 anomalies corresponded to events. If matching is relaxed to include the next day, five of the top 10 anomalies correspond to events. Using Tweet count time series, only one of the top 10 anomalies corresponds to an event. When including the next day, an additional event is detected. This demonstrates favorable performance when using motif time series in comparison to Tweet count time series, as well as compared to the base rate. This suggests a possible relationship between anomalous central network structure captured by motifs and the occurrence of major news events. This is additionally supported by the fact that Tweet counts do not directly capture specific network structures, but simply the volume of network conversation.

Discussion and Conclusion

In this paper, we propose a novel motif time series feature, grounded in social science theory, that provides a framework to study dynamic SBPM networks. To investigate this feature, we collected data from known SBPM campaigns as well as authentic data that are representative and share behavior with SBPM campaigns. Using MFDA, we can achieve comparable AUC scores to the highest performing alternatives. Our classification results support the hypothesis that there are distinct temporal and structural qualities to dynamic SBPM networks. Further taking advantage of our temporal feature, we perform anomaly detection, achieving superior performance in comparison to using Tweets counts. Specifically, our findings demonstrate a potential link between changes in central network structure captured by the motifs we use and the occurrence of newsworthy events. In conclusion, motif time series provide a promising framework to study SBPM, allowing for a variety of static and dynamic analyses.

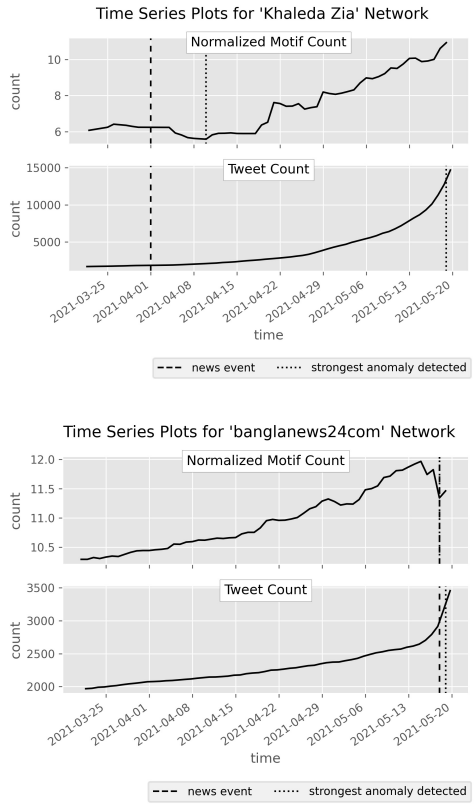

Figure 6: Time series plots showing an example of anomaly detection applied to two authentic topic networks. Each plot shows the raw motif count and Tweet count time series data. The date of an event pertaining to that topic network is marked, as well as the date of the strongest anomaly given the time series.

For further work, we first note that MFDA applies a spline model to each time series separately. A global model should be investigated to see if embeddings can be learned more efficiently. Second, implementing our approach on more network types would help bridge the comparison between our approach and others. Third, in our early analysis, change point detection did not prove useful with our datasets. In this setting, it may be necessary to use data with longer time scales or use different processing for motif time series data. Fourth, we only considered motifs that were abundant in our data. For less commonly present motifs, e.g. the 3-cycle, computing the ratio of multiple motifs together would allow them to be incorporated. Fifth, the network data we used had a relatively high variance in size (coefficient of variation of nodes equal to 1.62). We explored using datasets with a controlled number of users and Tweets, which in turn controls network size. In this restricted setting, we found that AUC scores increase overall across feature sets, except for static motifs. Potential improvements include using a refined normalization method, incorporating network size in all feature sets, or supplementing the data across all network scales. Sixth, understanding any potential biases from truncating

periods could be important future work. Lastly, there is additional interest in understanding the similarities between different campaigns, beyond whether they are SBPM or not. Clustering with motif time series allows for finer categorization of campaigns by their evolving network structure, which is useful to supplement understanding the different styles of SBPM campaigns that exist.

Ethical Statement

On a small scale, this paper will advance the study of state-sponsored influence online. More broadly, our work aims to help institutions analyze SBPM online to mitigate its influence. There are already efforts to understand and mitigate SBPM on social media, because of its detrimental effects on societies around the globe (Zannettou et al. 2019). Any method that contributes to this effort promotes authentic discourse online.

Twitter users agree to allow their data to be made publicly available when signing the terms of service. Twitter's algorithm for selecting data is not publicly known, so any biases in their algorithm may manifest in models trained on Twitter data. We intentionally develop and analyze a variety of datasets in an effort to have a better representation of the Twitter population.

This research could be used in both the public and private sectors to allow the development of more adaptive policies that address SBPM. Our approach is intended as a screening method, not as a standalone method, to help identify potential state-sponsored influence, and we believe it is most effectively incorporated into a system to avoid misidentification and false positives.

Acknowledgements

The authors would like to thank Clint Watts of Miburo Solutions for his collaboration and support. This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government.

References

Akoglu, L.; Tong, H.; and Koutra, D. 2015. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3): 626–688.

Alhazbi, S. 2020. Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter. *IEEE Access*, 8: 195132–195141.

Alizadeh, M.; Shapiro, J. N.; Buntain, C.; and Tucker, J. A. 2020. Content-based features predict social media influence operations. *Science Advances*, 6(30): eabb5824.

Badawy, A.; Addawood, A.; Lerman, K.; and Ferrara, E. 2019. Characterizing the 2016 Russian IRA influence campaign. *Social Network Analysis and Mining*, 9(1): 31.

Badawy, A.; Lerman, K.; and Ferrara, E. 2019. Who Falls for Online Political Manipulation? In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, 162–168. New York, NY, USA: Association for Computing Machinery.

Bail, C. A.; Guay, B.; Maloney, E.; Combs, A.; Hillygus, D. S.; Merhout, F.; Freelon, D.; and Volfovsky, A. 2020. Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proceedings of the National Academy of Sciences*, 117(1): 243–250.

Beskow, D. M.; and Carley, K. M. 2020. Characterization and Comparison of Russian and Chinese Disinformation Campaigns. In Shu, K.; Wang, S.; Lee, D.; and Liu, H., eds., *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, Lecture Notes in Social Networks, 63–81. Cham: Springer International Publishing.

Bhamidi, S.; Jin, J.; and Nobel, A. 2018. Change point detection in network models: Preferential attachment and long range dependence. *The Annals of Applied Probability*, 28(1): 35–78.

Box, G. E. P.; and Cox, D. R. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2): 211–252.

Bradshaw, S.; and Howard, P. 2017. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. *Oxford Internet Institute*, 2017.12: 1–37. Publisher: Oxford Internet Institute.

Bunke, H.; Dickinson, P. J.; Kraetzl, M.; and Wallis, W. D. 2007. *A Graph-Theoretic Approach to Enterprise Network Dynamics*. Progress in Computer Science and Applied Logic. Birkhauser Basel.

Cangea, C.; Velickovic, P.; Jovanovic, N.; Kipf, T.; and Lio, P. 2018. Towards Sparse Hierarchical Graph Classifiers. ArXiv preprint at http://arxiv.org/abs/1811.01287.

Chakrabarti, R.; and Berthon, P. 2012. Gift giving and social emotions: experience as content. *Journal of Public Affairs*, 12(2): 154–161.

Cheng, A.; and Dickinson, P. 2013. Using Scan-Statistical Correlations for Network Change Analysis. In Li, J.; Cao, L.; Wang, C.; Tan, K. C.; Liu, B.; Pei, J.; and Tseng, V. S., eds., *Trends and Applications in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science, 1–13. Berlin, Heidelberg: Springer.

Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; and Tesconi, M. 2017. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, 963–972.

Ferrara, E. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday*, 22.

Fire, M.; Katz, G.; and Elovici, Y. 2012. Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies. *Human Journal*, 1(1): 26–39.

Fisch, A.; Grose, D.; Bardwell, L.; Eckley, I.; and Fearnhead, P. 2020. anomaly: Detecting Anomalies in Data. ArXiv preprint at https://arxiv.org/abs/2010.09353.

Fisch, A. T. M.; Eckley, I. A.; and Fearnhead, P. 2018. A linear time method for the detection of point and collective anomalies. ArXiv preprint at http://arxiv.org/abs/1806.01947.

Friggeri, A.; Adamic, L.; Eckles, D.; and Cheng, J. 2014. Rumor Cascades. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1).

Ghanem, B.; Buscaldi, D.; and Rosso, P. 2019. TexTrolls: Identifying Russian Trolls on Twitter from a Textual Perspective. *ArXiv*, abs/1910.01340.

Gill, P.; Corner, E.; Conway, M.; Thornton, A.; Bloom, M.; and Horgan, J. 2017. Terrorist Use of the Internet by the Numbers. *Criminology & Public Policy*, 16(1): 99–117.

Grimme, C.; Assenmacher, D.; and Adam, L. 2018. Changing Perspectives: Is It Sufficient to Detect Social Bots? In *HCI*.

Guo, X.; and Vosoughi, S. 2020. Multi-modal Identification of State-Sponsored Propaganda on Social Media. *ArXiv*, abs/2012.13042.

Hamilton, W. L.; Ying, R.; and Leskovec, J. 2018. Representation Learning on Graphs: Methods and Applications. *ArXiv*, abs/1709.05584.

Hastie, T.; and Tibshirani, R. 1986. Generalized Additive Models. *Statistical Science*, 1(3): 297–310.

Heard, N. A.; Weston, D. J.; Platanioti, K.; and Hand, D. J. 2010. Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics*, 4(2): 645–662.

Hendrickson, S.; Kolb, J.; Lehman, B.; and Montague, J. 2015. Trend Detection in Social Data. https://blog.twitter.com/2015/trend-detection-social-data. Accessed: 2023-04-12.

Im, J.; Chandrasekharan, E.; Sargent, J.; Lighthammer, P.; Denby, T.; Bhargava, A.; Hemphill, L.; Jurgens, D.; and Gilbert, E. 2020. Still out there: Modeling and Identifying Russian Troll Accounts on Twitter. In *12th ACM Conference on Web Science*, WebSci '20, 1–10. New York, NY, USA: Association for Computing Machinery.

Johnston, R.; Watts, C.; and Younce, B. 2020. Social Sifter. https://symposium.ncsu-las.net/2020/influence.html. Accessed: 2023-04-12.

Juul, J. L.; and Ugander, J. 2021. Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences of the United States of America*, 118(46).

Kendrick, L.; Musial, K.; and Gabrys, B. 2018. Change point detection in social networks–Critical review with experiments. *Computer Science Review*, 29: 1–13.

Knobloch-Westerwick, S.; Mothes, C.; and Polavin, N. 2020. Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. *Communication Research*, 47(1): 104–124.

Lauschke, C.; and Ntoutsi, E. 2012. Monitoring User Evolution in Twitter. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 972–977.

Linvill, D. L.; and Warren, P. L. 2020. Troll Factories: Manufacturing Specialized Disinformation on Twitter. *Political Communication*, 37(4): 447–467.

Luceri, L.; Giordano, S.; and Ferrara, E. 2020. Detecting Troll Behavior via Inverse Reinforcement Learning: A Case Study of Russian Trolls in the 2016 US Election. *Proceedings of the International AAAI Conference on Web and Social Media*, 14: 417–427.

Mach, Z. 1993. *Symbols, conflict, and identity: essays in political anthropology*. SUNY series in anthropological studies of contemporary issues. Albany: State University of New York Press.

Martino, G. D. S.; Cresci, S.; Barron-Cedeno, A.; Yu, S.; Pietro, R. D.; and Nakov, P. 2020. A Survey on Computational Propaganda Detection. In Bessiere, C., ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, 4826–4832. International Joint Conferences on Artificial Intelligence Organization.

Marx, B. D.; and Eilers, P. H. 1999. Generalized Linear Regression on Sampled Signals and Curves: A P-Spline Approach. *Technometrics*, 41(1): 1–13.

Maugis, P.-A. G.; Olhede, S. C.; and Wolfe, P. J. 2017. Topology reveals universal features for network comparison. *ArXiv*, abs/1705.05677.

Mauss, M. 2002. *The Gift: The Form and Reason for Exchange in Archaic Societies*. Routledge.

Miller, B. A.; Arcolano, N.; and Bliss, N. T. 2013. Efficient anomaly detection in dynamic, attributed graphs: Emerging phenomena and big data. In *2013 IEEE International Conference on Intelligence and Security Informatics*, 179–184.

Moghaddam, F. M. 2005. The Staircase to Terrorism: A Psychological Exploration. *American Psychologist*, 60(2): 161–169.

Mongiovi, M.; Bogdanov, P.; Ranca, R.; Papalexakis, E. E.; Faloutsos, C.; and Singh, A. K. 2013. NetSpot: Spotting Significant Anomalous Regions on Dynamic Networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining (SDM)*, Proceedings, 28–36. Society for Industrial and Applied Mathematics.

Neil, J.; Hash, C.; Brugh, A.; Fisk, M.; and Storlie, C. B. 2013. Scan Statistics for the Online Detection of Locally Anomalous Subgraphs. *Technometrics*, 55(4): 403–414.

Nye, J. S. 2002. The Information Revolution and American Soft Power. *Asia-Pacific Review*, 9(1): 60–76.

Pacheco, D.; Hui, P.-M.; Torres-Lugo, C.; Truong, B. T.; Flammini, A.; and Menczer, F. 2021. Uncovering coordinated networks on social media: Methods and case studies. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1): 455–466.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.;

Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830.

Ranshous, S.; Shen, S.; Koutra, D.; Harenberg, S.; Faloutsos, C.; and Samatova, N. F. 2015. Anomaly detection in dynamic networks: a survey. *WIREs Computational Statistics*, 7(3): 223–247.

Rossi, R. A.; Gallagher, B.; Neville, J.; and Henderson, K. 2013. Modeling dynamic behavior in large evolving graphs. In *Proceedings of the sixth ACM international conference on Web search and data mining*, WSDM '13, 667–676. New York, NY, USA: Association for Computing Machinery.

Savage, D.; Zhang, X.; Yu, X.; Chou, P.; and Wang, Q. 2014. Anomaly detection in online social networks. *Social Networks*, 39: 62–70.

Savitzky, A.; and Golay, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8): 1627–1639.

Serven, D.; Brummitt, C.; Abedi, H.; and Hlink. 2018. Dswah/Pygam: V0.8.0. https://zenodo.org/record/1476122#.ZDdhqC_MIeY. Accessed: 2023-04-12.

Skageby, J. 2010. Gift-Giving as a Conceptual Framework: Framing Social Behavior in Online Networks. *Journal of Information Technology*, 25(2): 170–177.

Stewart, L. G.; Arif, A.; and Starbird, K. 2018. Examining Trolls and Polarization with a Retweet Network. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 6. ACM.

Takahashi, T.; Tomioka, R.; and Yamanishi, K. 2014. Discovering Emerging Topics in Social Streams via Link-Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, 26(1): 120–130.

Thaler, R. H. 2016. Behavioral Economics: Past, Present, and Future. *American Economic Review*, 106(7): 1577–1600.

Tu, K.; Li, J.; Towsley, D.; Braines, D.; and Turner, L. D. 2019. Gl2vec: Learning Feature Representation Using Graphlets for Directed Networks. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '19, 216–221. New York, NY, USA: Association for Computing Machinery.

Vargas, L.; Emami, P.; and Traynor, P. 2020. On the Detection of Disinformation Campaign Activity with Network Analysis. In *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*, CCSW'20, 133–146. New York, NY, USA: Association for Computing Machinery.

Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; and van Mulbregt, P. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3).

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science*, 359(6380): 1146–1151.

Wang, J.-L.; Chiou, J.-M.; and Müller, H.-G. 2016. Functional data analysis. *Annual Review of Statistics and Its Application*, 3: 257–295.

Wolf, E. R. 1999. *Envisioning Power: Ideologies of Dominance and Crisis*. University of California Press.

Woolley, S. C.; and Howard, P. 2017. Computational propaganda worldwide: Executive summary. Technical report, Oxford Internet Institute, Oxford, UK.

Woolley, S. C.; and Howard, P. N. 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press.

Yu, R.; He, X.; and Liu, Y. 2015. GLAD: Group Anomaly Detection in Social Media Analysis. *ACM Transactions on Knowledge Discovery from Data*, 10(2): 18:1–18:22.

Yu, R.; Qiu, H.; Wen, Z.; Lin, C.; and Liu, Y. 2016. A Survey on Social Media Anomaly Detection. *ACM SIGKDD Explorations Newsletter*, 18(1): 1–14.

Zannettou, S.; Caulfield, T.; Setzer, W.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019. Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, 353–362. New York, NY, USA: Association for Computing Machinery.

Zhang, J.; Wei, Z.; Yan, Z.; and Pani, A. 2015. Collaborated Online Change-Point Detection in Sparse Time Series for Online Advertising. In *2015 IEEE International Conference on Data Mining*, 1099–1104.