UNSUPERVISED DETECTION OF COORDINATED FAKE FOLLOWERS ON SOCIAL MEDIA

by
YASSER ZOUZOU

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
July 2024

ABSTRACT

UNSUPERVISED DETECTION OF COORDINATED FAKE FOLLOWERS ON SOCIAL MEDIA

YASSER ZOUZOU

DATA SCIENCE M.S. THESIS, JULY 2024

Thesis Supervisor: Prof. ONUR VAROL

Keywords: computational social science, fake-followers, bots, online coordinated activities

Social media accounts are known to have automated accounts, referred to as social bots, among their active users. While these accounts are not necessarily harmful, they are widely used to manipulate engagement metrics and in coordinated disinformation campaigns. In this work, we propose a novel unsupervised approach to detect a subset of social bots, known as fake followers, which are used to deceitfully amplify online popularity of users. Our method identifies fake followers by detecting anomalous temporal following patterns within the followers of a social media account. Furthermore, we use our method to investigate the prevalence of anomalous followers in the Turkish political Twitter network (later rebranded as X). In addition to detection, we also demonstrated that groups of anomalous followers may act in coordination across several accounts in the same network. Our results show that the proposed framework can be used to investigate large-scale coordinated manipulation campaigns on social media platforms.

ÖZET

SOSYAL MEDYADA KOORDINELI SAHTE TAKIPÇI DENETIMSIZ TESPITI

YASSER ZOUZOU

VERİ BİLİMİ YÜKSEK LİSANS TEZİ, TEMMUZ 2024

Tez Danışmanı: Dr. Öğr. Üyesi ONUR VAROL

Anahtar Kelimeler: Hesaplamalı sosyal bilimleri, sahte takipçi, bot, çevrimiçi koordineli faaliyet

Sosyal medya etkin kullanıcıları arasında "sosyal botlar" olarak adlandırılan otomatik hesapların bulunduğu bilinmektedir. Bu hesaplar zararlı olmak zorunda olmasa da, genellikle etkileşim ölçütlerini manipüle etmek ve koordineli yanıltıcı bilgi kampanyalarında kullanılmaktadır. Bu çalışmada, kullanıcıların çevrimiçi popülerliklerini hileli olarak artırmak için kullanılan "sahte takipçiler" olarak bilinen bir sosyal bot alt kümesini tespit etmek için yeni bir denetimsiz yaklaşım öneriyoruz. Yöntemimiz, bir sosyal medya hesabının takipçileri arasındaki anormal zamansal takip etme düzenleri tespit ederek sahte takipçileri belirler. Ayrıca, yöntemimizi Türk siyasi Twitter (daha sonra X olarak adlandırılan) ağındaki anormal takipçilerin yaygınlığını incelemek için kullandık. Bot tespitinin yanı sıra, aynı ağdaki çeşitli hesaplar üzerinde koordinasyon içinde hareket edebilecek anormal takipçi gruplarının varlığını da gösterdik. Sonuçlarımız, önerdiğimiz metodun sosyal medya platformlarındaki büyük ölçekli koordineli manipülasyon kampanyalarını incelemek için kullanılabileceğini göstermektedir.

ACKNOWLEDGEMENTS

First and foremost I am thankful to Allah for without His graces and blessings I would not have been here.

I am grateful to Onur Hocam for his continuous support and for all the knowledge, research skills, and analytical skills I was privileged to learn from him throughout this master's program.

I would also like to give special thanks to my family for their unconditional support, without you none of this would have been possible.

Finally, I want to Sabanci University for their generous scholarship and TÜBİTAK for supporting our research through the project grant 222N311.

To my beloved family..

TABLE OF CONTENTS

LIST OF TABLES ..... **x**

LIST OF FIGURES ..... **xi**

1. INTRODUCTION **1**
1.1. General Flow of the Thesis 2

2. RELATED WORK **3**
2.1. Misinformation on Social Media 3
2.2. Social Bots 4
2.3. Bot Detection 5
2.4. Detecting Coordinated Activities 5

3. Data Description **7**
3.1. Dribbble dataset 7
3.2. Twitter dataset 7

4. Methodology **9**
4.1. Follower Map 9
4.2. Data processing 11
4.3. Synthetic Fake-Follower Data 12
4.4. Unsupervised Anomalous Follower Detection 14
4.4.1. Feature Engineering 14
4.4.2. Isolation Forest 16
4.4.3. Local Outlier Factor 17
4.4.4. ECOD 17
4.4.5. Gen2Out 18
4.4.6. Sliding Histogram 19
4.5. Follow-Time Estimation 21

5. Results **22**

5.1. Results on synthetic data 22
5.2. Results on real data 24
5.2.1. Retrieving users with anomalous followers 24
5.2.2. Identifying individual anomalous accounts 25
5.2.3. Exploring anomalous follower group behavior 28

6. Discussion 31
6.1. Anomalous follower detection 31
6.2. The case of Turkish political Twitter 32
6.3. Limitations 32

7. Conclusion 34

BIBLIOGRAPHY 35

APPENDIX A 40

LIST OF TABLES

Table 4.1. Parameter values for synthetic follower generation. 14
Table 4.2. Unsupervised anomaly detection feature definitions 16

Table 5.1. Results on Dribbble dataset. Area under ROC curve, average precision, and precision at 50 mean (std) values for all methods using different window sizes. 23

Table A.1. Internet Archive Wayback Machine links to the anomalous follower profiles presented in Fig. A.1-A.3. 41

LIST OF FIGURES

Figure 3.1. Follower numbers in the Dribbble (a) and Twitter (b) datasets 8

Figure 4.1. Follower map. (a) Follower map from the New York Times investigation with the fake followers shown in red Confessore, Dance, Harris & Hansen (2018) (b) Follower map from Varol & Uluturk (2020) (c) Twitter user follower map having normal followers (d) Twitter user follower map having anomalous followers. 10

Figure 4.2. A follower map of the first 15,000 followers of a Twitter user. The anomalous following patterns are highlighted in yellow. The vertical dashed lines mark the beginning of each year based on the estimated following times. 10

Figure 4.3. Followers with erroneous creation dates (marked by redboxes) in the Dribbble (a-b) and Twitter (c-d) datasets. 11

Figure 4.4. The two types of artificial anomalous followers and their corresponding parameters 13

Figure 4.5. Profile creation date lower and upper bounds, and centered window around the corresponding follower for which features are computed 15

Figure 4.6. Illustration of the Sliding Histogram (a) A follower map with inserted synthetic irregular followers (orange) showing all sliding windows (light gray) with two of them highlighted in black and orange. (b) The histograms corresponding to the two highlighted windows in the follower map. Window 1 only includes normal followers and Window two includes anomalous followers. The numbers are the count of followers that fall within each bin. (c) All histograms plotted together as line plots, with the black and orange lines corresponding to the black and orange windows above. (d) A zoom in on bin No. 5 showing the median and interquartile range (IQR) of all histograms at this bin. 20

Figure 4.7. Mean follow time estimation error. Each point represents the error between the estimated follow time and the ground truth averaged across all followers of one Dribbble user. The mean error is less than one day for users with more than 10,000 followers. 21

Figure 5.1. Heatmap of AUC (top) and AP (bottom) of the ECOD method for all the synthetic cases plotted by the ratio of anomalous followers to the total number of followers in each case. The plotted values correspond to the window size 101. 23
Figure 5.2. Heatmap of AUC (top) and AP (bottom) of the SH method for all the synthetic cases plotted by the ratio of anomalous followers to the total number of followers in each case. The plotted values correspond to the window size 101. 24
Figure 5.3. Retrieving users with anomalous followers Follower maps of the 9 Twitter accounts with the highest average anomaly score across all of their followers. The colors represent the average anomaly scores of all followers that fall in each bin (cell) of the heat map. 25
Figure 5.4. Follower maps of 4 popular Twitter accounts (>500k followers) with anomalous followers. The sub figures under each follower map are a zoom-in on the parts marked by a red box on the main follower map. 26
Figure 5.5. Follower maps of 4 Twitter accounts with obvious anomalous followers. The sub figures under each follower map are a zoom-in on the parts marked by a red box on the main follower map. 26
Figure 5.6. Detailed analysis of anomalous followers. User A: Anomalous followers have high bot scores. User B: Anomalous followers have low bot scores. Anomalous regions are zoomed in for User A (b,c) and User B (f,g). Profile statistics for regular and all followers are also compared for these users in subplots (d) and (e). 27
Figure 5.7. Coordinated behavior of anomalous followers. Follow times (top) and anomaly scores (bottom) of the shared anomalous followers (red) and the shared non-anomalous followers (gray) across 13 users that are followed by the same batch of anomalous followers shown in Fig.5.6(f). 28

Figure 5.8. Similarity network based on the shared anomalous followers. The full network is shown in the middle of the figure, where nodes are colored based on communities and sized based on their degrees. The two communities with the highest pairwise average anomaly scores are highlighted and shown in detail along with the follower maps of one edge in each community. 30

Figure 6.1. Users that have a high ratio of anomalous followers. 33

Figure A.1. Sample of anomalous followers of @nurettincanikli. The follower map of this user is shown in Fig. 5.7a 40
Figure A.2. Sample of anomalous followers of @yigitbulutt. The follower map of this user is shown in Fig. 5.7e 40
Figure A.3. Sample of anomalous followers of @matillakaya. The follower map of this user is shown in Fig. A.6 41
Figure A.4. Shared anomalous followers. Follower maps of the 6 user pairs corresponding to the highest similarity scores in our dataset. 42
Figure A.5. Network of shared anomalous followers. Node colors represent community membership and node sizes are scaled by node degrees ... 43
Figure A.6. Follower maps of user pairs corresponding to 4 edges in one of the communities of the shared anomalous followers network. The follow pattern of these anomalies are similar across several users. 43

1. INTRODUCTION

Social media usage has largely increased since the early 2000s Perrin (2015), gaining popularity amongst individuals from all ages and socioeconomic statuses Auxier & Anderson (2021). While social media platforms were initially spaces in which individuals share their experiences and thoughts on various topics, they have been recently used by official institutions and politicians as a means of communication with the public Jungherr (2014). In a recent survey conducted by the Pew Research Center in the US, 33% of the respondents cited that they "Sometimes" used social media as a source for news, and 17% "Often" did so Liedke & Matsa (2022). This increasing usage of social media as a venue for news dissemination has rendered it a favourable place for politicians to run their political campaigns and for researchers to track public opinion Anstead & O'Loughlin (2015); DiGrazia, McKelvey, Bollen & Rojas (2013); Jungherr (2016); Metaxas & Mustafaraj (2012). As a natural consequence of this, social media became the target of misinformation campaigns to influence public opinion, undermine trust in institutions, and impact democratic processes Deb, Luceri, Badaway & Ferrara (2019); Faris, Roberts, Etling, Bourassa, Zuckerman & Benkler (2017); Morgan (2018); Ratkiewicz, Conover, Meiss, Gonçalves, Flammini & Menczer (2011). In order to enable large scale manipulation campaigns, automated social media accounts known as *social bots* have been widely used Bruno, Lambiotte & Saracco (2022); Cresci, Di Pietro, Petrocchi, Spognardi & Tesconi (2017a); Ferrara, Varol, Davis, Menczer & Flammini (2016); Himelein-Wachowiak, Giorgi, Devoto, Rahman, Ungar, Schwartz, Epstein, Leggio & Curtis (2021); Mendoza, Tesconi & Cresci (2020); Shao, Ciampaglia, Varol, Flammini & Menczer (2017); Shao, Ciampaglia, Varol, Yang, Flammini & Menczer (2018). The rising prevalence of social bots on online platforms has made bot detection a focal point in research Cresci, Di Pietro, Petrocchi, Spognardi & Tesconi (2015); Ding & Chen (2023); Liu, Tan, Wang, Feng, Zheng & Luo (2023); Mazza, Cresci, Avvenuti, Quattrociocchi & Tesconi (2019); Takacs & McCulloh (2019); Yang, Varol, Davis, Ferrara, Flammini & Menczer (2019).

In this research, we introduced a novel unsupervised method to detect a type of

coordinated bots on social media that has not been specifically addressed by detection methods before Zouzou & Varol (2023). By looking at the temporal following patterns of a user's followers, we detected anomalous patterns that corresponded to automated fake followers. In particular, we detected followers that had similar profile creation dates and followed users almost simultaneously. While coordination in account activity does not necessarily coincide with automation Nizzoli, Tardelli, Avvenuti, Cresci & Tesconi (2021); Pacheco, Hui, Torres-Lugo, Truong, Flammini & Menczer (2021), accounts created on similar dates and following users successively are more likely to be automated Bellutta & Carley (2023); Confessore et al. (2018); Varol & Uluturk (2020). Indeed, this is intuitive because humans may engage in similar social media activities, such as posting about hot topics. However, there is no reason for accounts created on similar dates to follow the same users simultaneously. Furthermore, we conducted a case-study on the Turkish political Twitter network in which we identified fake followers and analyzed their coordinated behavior across different politician accounts.

Research objectives:

- To propose a novel unsupervised method for detecting fake followers based on temporal following patterns
- To provide insights on coordinated manipulation campaigns in Turkish political Twitter

1.1 General Flow of the Thesis

Chapter 2 provides a literature review of online misinformation campaigns, coordinated activities on social media, and bot detection methods. Chapter 3 presents our methodology and describes the datasets we used in this study. Results and main finding are presented in Chapter 4, and the discussion and conclusion is left for Chapter 5.

2. RELATED WORK

2.1 Misinformation on Social Media

The expanding outreach of social media Auxier & Anderson (2021); Perrin (2015), along with the speed of information diffusion on its platforms has made it an ideal setting for misinformation spreading Muhammed T & Mathew (2022). Misinformation can be defined as fake or inaccurate information that is spread intentionally or unintentionally Wu, Morstatter, Carley & Liu (2019). Misinformation on social media is frequently observed in the context of health Wang, McKee, Torbica & Stuckler (2019), conspiracy theories Bessi, Coletto, Davidescu, Scala, Caldarelli & Quattrociocchi (2015); Cinelli, Etta, Avalle, Quattrociocchi, Di Marco, Valensise, Galeazzi & Quattrociocchi (2022), and politics Morgan (2018); Tucker, Guess, Barberá, Vaccari, Siegel, Sanovich, Stukal & Nyhan (2018). The recent COVID 19 outbreak and the consequent debate on vaccine and mask regulations was a vivid example of the danger of misinformation dissemination on social media Seckin, Atalay, Otenen, Duygu & Varol (2024); Singh, Lima, Cha, Cha, Kulshrestha, Ahn & Varol (2022). In the context of politics, numerous research has been done to highlight the role of misinformation in opinion manipulation Keller, Schoch, Stier & Yang (2020); Ratkiewicz et al. (2011). With the recent advancements in the field of natural language processing, large language models pre-trained on social media text have been developed, which facilitates analyzing shared textual content on social media Najafi & Varol (2024a,2); Qudar & Mago (2020). A main driver of misinformation spreading on social media is the use of automated accounts, also known as *social bots* Shao et al. (2018).

2.2 Social Bots

Social bots are automated computer programs that interact with humans and share information on social media platforms Ferrara et al. (2016). Social bots can be helpful or harmful. Helpful bots include ones that post news updates and weather forecasts, or ones that automatically reply to users Boshmaf, Muslukhov, Beznosov & Ripeanu (2013); Varol, Ferrara, Davis, Menczer & Flammini (2017). Twitter, which was recently rebranded as X, allows bots to be run through the official API but requires them to be self-declared bots Alkulaib, Zhang, Sun & Lu (2022); Yang et al. (2019). On the other hand, harmful social bots can have different types; fake followers that inflate follower numbers to provide an illusion of popularity, spam bots that share and engage with posts in high frequencies to manipulate engagement metrics and flood social media with certain information, and human behavior emulators that promote certain propaganda, rumors, and conspiracy theories Cresci et al. (2017a); Ferrara, Wang, Varol, Flammini & Galstyan (2016); Hristakieva, Cresci, Da San Martino, Conti & Nakov (2022); Pierri, Luceri, Jindal & Ferrara (2023); Ratkiewicz et al. (2011). In a study on the role of social bots in the 2016 US presidential elections, it was found that social bots accounted for about one fifth of the political discourse during one month prior to the elections Bessi & Ferrara (2016). In another study investigating a Syrian network of bots, the bots' main role was "smoke screening" by posting irrelevant content and using hashtags related to the Syrian civil war in an attempt to divert attention from the content of the original hashtags Abokhodair, Yoo & McDonald (2015). *Astroturfing* is another way in which social bots can have a malicious role on social media. Astroturfing refers to coordinated social bots that are centrally directed to imitate human behavior and create an illusion of grassroots activism Keller, Schoch, Stier & Yang (2017); Keller et al. (2020); Schoch, Keller, Stier & Yang (2022); Zhang, Carpenter & Ko (2013). The adverse role of bots on social media has rendered detecting them a primary area of research.

2.3 Bot Detection

Cresci provided a comprehensive review of bot detection methods that were proposed throughout the last decade Cresci (2020). From an algorithmic perspective, the methods for detecting bots are divided into supervised and unsupervised approaches. Supervised detection methods usually consist of classifiers trained on features extracted from account metadata, network properties, textual features from the shared posts, temporal features, or a mixture of these features Ding & Chen (2023); Liu et al. (2023); Sayyadiharikandeh, Varol, Yang, Flammini & Menczer (2020); Varol, Davis, Menczer & Flammini (2018). Labeled datasets of human and bot accounts, which are limited in availability and insufficient to capture the types and evolution of bots, constitute the main shortcoming of supervised detection methods Echeverrï£¡ a, De Cristofaro, Kourtellis, Leontiadis, Stringhini & Zhou (2018). On the other hand, unsupervised methods rely on the assumption that bots have a similar behavior among themselves, which is different from human behavior on social media platforms. Therefore, by clustering users based on a predefined set of features, clusters that have suspicious properties can be identified as bots Mannocci, Cresci, Monreale, Vakali & Tesconi (2022); Mazza et al. (2019). While unsupervised methods are not prone to the bias introduced by labeled datasets, they are still biased to the presumptions that define what constitutes anomalous or malicious behaviors. Finally, semi-supervised methods, in which a small part of the dataset is labeled, have also been used in the bot detection literature Jia, Wang & Gong (2017); Mendoza et al. (2020). These methods generally rely on a representing user interactions in networks and identifying the users that are close to labeled bot accounts as suspicious accounts.

2.4 Detecting Coordinated Activities

Methods for detecting coordinated activities on social media are also essential for identifying online manipulation campaigns. Coordination detection methods rely on defining a similarity measure between users and identifying groups of users that are unexpectedly similar to each other. The similarity measures used in the literature include similarity based on the shared content Nizzoli et al. (2021); Pacheco et al. (2021), temporal correlation in activities on social media Chavoshi, Hamooni & Mueen (2016); Cresci, Di Pietro, Petrocchi, Spognardi & Tesconi (2017b); Pacheco et al. (2021); Sharma, Zhang, Ferrara & Liu (2021), identity Pacheco et al. (2021), or a combination of several measures Magelinski, Ng & Carley (2022); Pacheco

et al. (2021); Weber & Neumann (2021). These methods assume that user activities on social media are mostly independent and a significant interdependence in their activities indicates coordination. A recent study on coordinated online influence campaigns indicated that groups of followers created in short periods exhibited similar behavior amongst themselves and were more likely to be bots Bellutta & Carley (2023). Bursts of account creations were also observed around the dates of major political events in the US Takacs & McCulloh (2019). Furthermore, a New York Times investigation that tracked fake accounts sold in bulk as fake followers showed that these accounts tend to have similar creation dates and follow the target user successively Confessore et al. (2018). A subsequent study on journalists on Twitter identified similar patterns in fake followers which were used to increase online popularity and manipulate the online perception of journalist accounts Varol & Uluturk (2020). The aforementioned studies show that similarities in creation dates and follow times strongly indicate coordinated activity and possible automation. However, there are no detection methods that specifically address this type of coordination.

3. Data Description

3.1 Dribbble dataset

Dribbble is a platform digital designers use to share and promote their work. Due to the professional nature of the Dribbble dataset, we assumed that it would be less polluted by automated fake followers than Twitter data. This makes it a suitable dataset of "normal followers" to start with, into which we can insert synthetic anomalous followers to test different detection methods. Additionally, since the Dribbble platform provides the time each follower followed a certain user, this dataset served as a ground truth dataset to evaluate the follow-time estimation algorithm we used in this study. The Dribbble dataset comprises profile information and follow-times of the followers of 2,834 users. The collected users had between 1,000-110,000. The distribution of follower counts is shown in Fig. 3.1 (a). This dataset was used for (i) the creation of a synthetic dataset containing anomalous followers as described in Section 4.3 (ii) the evaluation of the follow-time estimation algorithm.

3.2 Twitter dataset

The Twitter dataset used in this study comprises Twitter accounts of Turkish politicians and media outlets and their corresponding follower profile information. This dataset is part of the #Secim2023 dataset Najafi, Mugurtay, Zouzou, Demirci, Demirkiran, Karadeniz & Varol (2024). The followers of each Twitter account are available as an ordered list of user IDs starting from the most recent follower to the oldest follower. We filter the dataset to include only users with more than 1,000

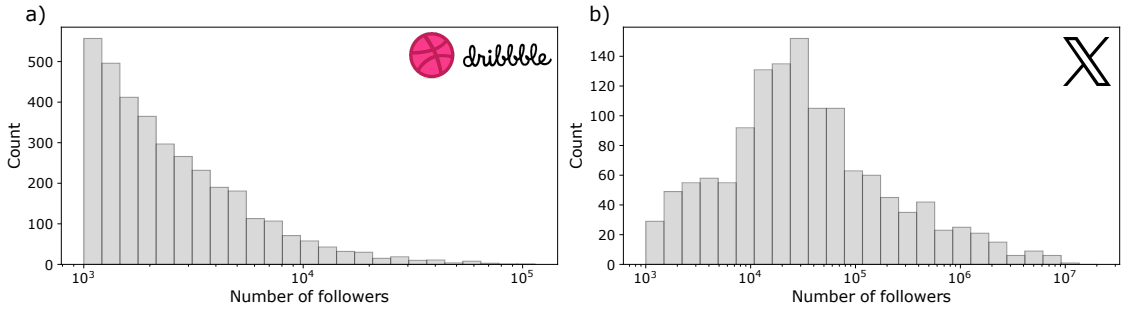

Figure 3.1 Follower numbers in the Dribbble (a) and Twitter (b) datasets

followers, resulting in a total of 1,318 accounts. The distribution of follower counts is shown in Fig. 3.1 (b). This dataset was used to explore the anomalous followers detected by our suggested method in the Turkish political Twitter circle.

4. Methodology

4.1 Follower Map

In an investigation conducted by New York Times, fake followers sold in bulk were found to be groups of accounts that were created in a small period and followed the user consecutively Confessore et al. (2018). This finding was supported by a recent study in which groups of Twitter accounts created in short periods were found to have a higher likelihood of being bots Bellutta & Carley (2023). These groups of followers can be visually distinguished on a plot that shows the followers of a user as a scatter plot, with the x-axis representing the follower ranks and the y-axis representing profile creation dates (Fig 4.1). We use the term *Follower Map* to refer to this type of graph. Fig. 4.2 shows an example of a follower map that has normal and anomalous followers. There is a clear difference between the distribution of followers in the anomalous following pattern zones and the normal ones. Furthermore, as seen in Figures 4.1 and 4.2, the follower map of each user has an increasing upper bound. This upper bound represents the profile creation date of the most recently created profile that has followed the user up to each follower rank. For each follower in a follower map, the value of the upper bound at his/her rank represents the minimum possible follow time of that follower, because each follower has certainly followed the user after the creation date of all previous follower profiles. In fact, we use the upper bound of the follower map to estimate the follow times of each follower, which is not provided by Twitter, based on the algorithm defined in Section 4.5. It can be seen that the upper bound remains almost horizontal in the anomalous zones, indicating that the fake followers follow the user almost simultaneously.The follower map is used throughout the study to show the fake followers of different Twitter users. For users with large numbers of followers, the follower map is plotted as a heat map instead of a scatter plot for better interpretability.

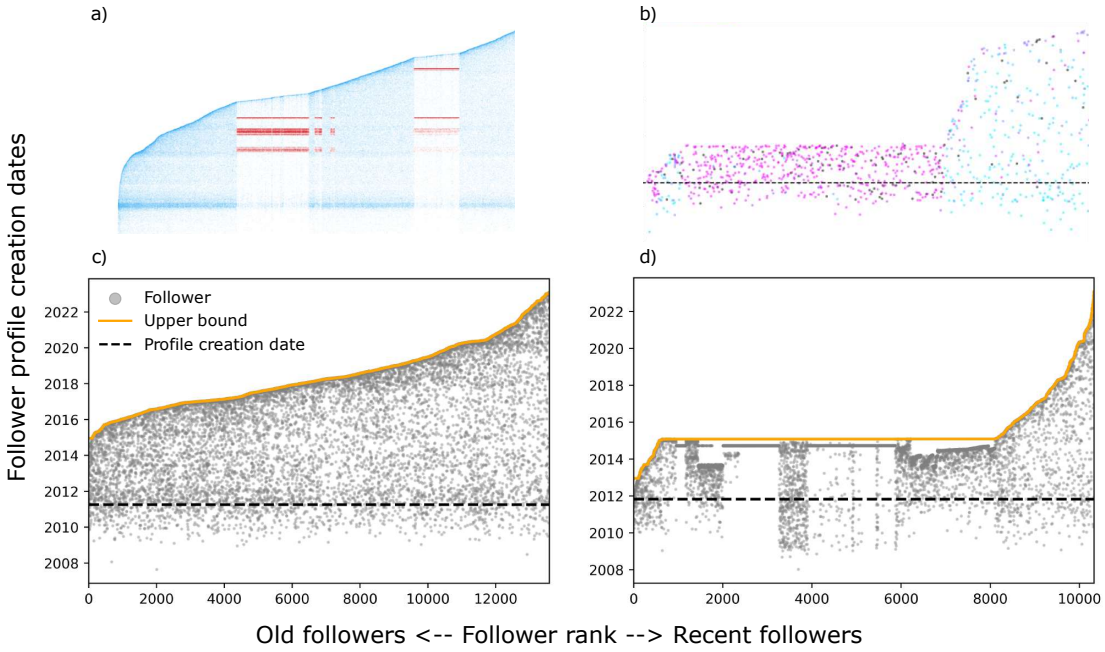

Figure 4.1 Follower map. (a) Follower map from the New York Times investigation with the fake followers shown in red Confessore et al. (2018) (b) Follower map from Varol & Uluturk (2020) (c) Twitter user follower map having normal followers (d) Twitter user follower map having anomalous followers.

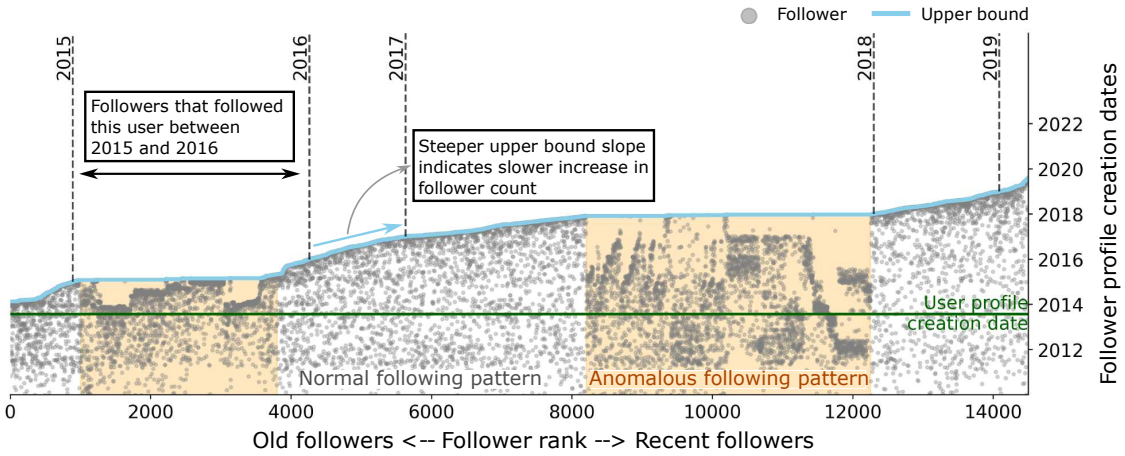

Figure 4.2 A follower map of the first 15,000 followers of a Twitter user. The anomalous following patterns are highlighted in yellow. The vertical dashed lines mark the beginning of each year based on the estimated following times.

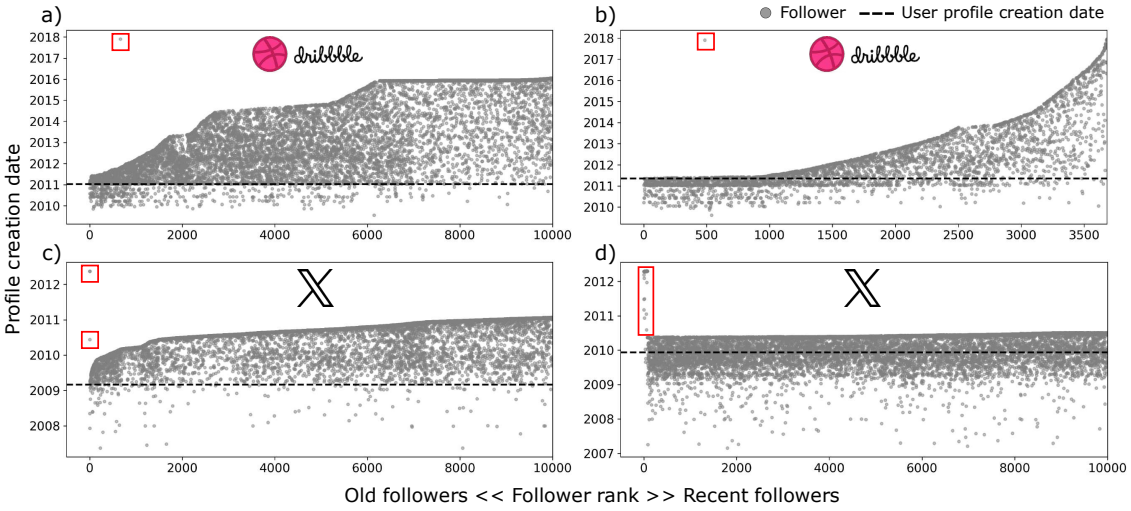

Figure 4.3 Followers with erroneous creation dates (marked by redboxes) in the Dribbble (a-b) and Twitter (c-d) datasets.

4.2 Data processing

In the Dribbble and Twitter datasets, we observed some followers that had unreasonable profile creation dates. These followers either had a creation date in 1970 (corresponding to 0 POSIX timestamp) or a creation date that lies way above the upper bound at the follower's rank (Fig. 4.3). Keeping these followers in the data results in a false shift in the lower and upper bounds of the profile creation dates, which affects the detection methods and the follow-time estimation algorithm. To identify users that have such followers, we used different approaches in the Dribbble and Twitter datasets. For the former, we identified the users that had an abnormally high average follow-time estimation error and then manually removed the erroneous followers. As for the Twitter dataset, which lacks ground truth follow times, we first identified users that had a sudden jump in the upper bound that exceeds a predefined threshold, then manually observed the followers of these users (90 users) and eliminated the invalid followers. In total, there were 2 Dribbble users and 20 Twitter users who had followers with incorrect profile creation dates. It is worth noting that all of the 20 Twitter users were created before 2012 and the erroneous followers were amongst the first followers.

4.3 Synthetic Fake-Follower Data

Due to the lack of a labeled dataset containing the type of anomalous followers we aim to detect in this study, we created a synthetic dataset to experiment with different detection methods before detecting anomalous followers in the real Twitter dataset. The synthetic dataset was generated by inserting artificial anomalous followers in a dataset of "normal followers". We used the Dribbble dataset, described in section 3.1, as our "normal follower" dataset since it is less susceptible to follower count manipulation campaigns than Twitter. As for the artificial anomalous followers, we generated two types of synthetic followers that simulate the temporal follow-pattern that has been observed in earlier studies Confessore et al. (2018); Varol & Uluturk (2020). Type 1 followers simulate a group of accounts that were created within a short period in the past and then consecutively follow the user. Type 2 followers simulate a collection of accounts that follow the user successively and right after their creation. The two types of artificial followers are shown in Fig. 4.4.

Since our detection approach only relies on the order of followers and their profile creation dates, generating artificial followers involves generating a list of follower user IDs and their corresponding profile creation dates. Subsequently, the generated list of followers is inserted at a certain position in the list of "normal followers" of a certain Dribbble account. For Type 1, we sampled the creation dates of a group of N_1 artificial followers from a normal distribution $\mathcal{N}(t_0, \sigma)$. t_0 was randomly sampled between the lower and upper bound of profile creation dates at the follower rank in which the artificial followers are to be inserted. The insertion rank was randomly chosen between the 40th and 60th percentiles of the follower ranks. The σ value determines the width of the time window in which the artificial follower batch was created, with lower values representing narrower time windows. For Type 2 synthetic followers, we duplicated the most recent N_{recent} followers that are on the upper bound of the profile creation dates of the Dribbble user's normal followers. Each of the N_{recent} followers was duplicated $N_{duplicate}$ times, which accounts for a total of $N_2 = N_{recent} * N_{duplicate}$ Type 2 followers.

We used the values for N_1, σ, N_2, and $N_{duplicate}$ depicted in Table 4.1 to create different permutations of synthetic anomalous follower data. We generated synthetic follower datasets with Type 1 only, Type 2 only, and a combined scenario with both types, in which we included the same number of each type. For each user in the Dribbble dataset, we inserted artificial anomalies using each possible permutation of the parameters in Table 4.1 in addition to the combined scenario, resulting in a total of $55 \times 2,834$ synthetic datasets.

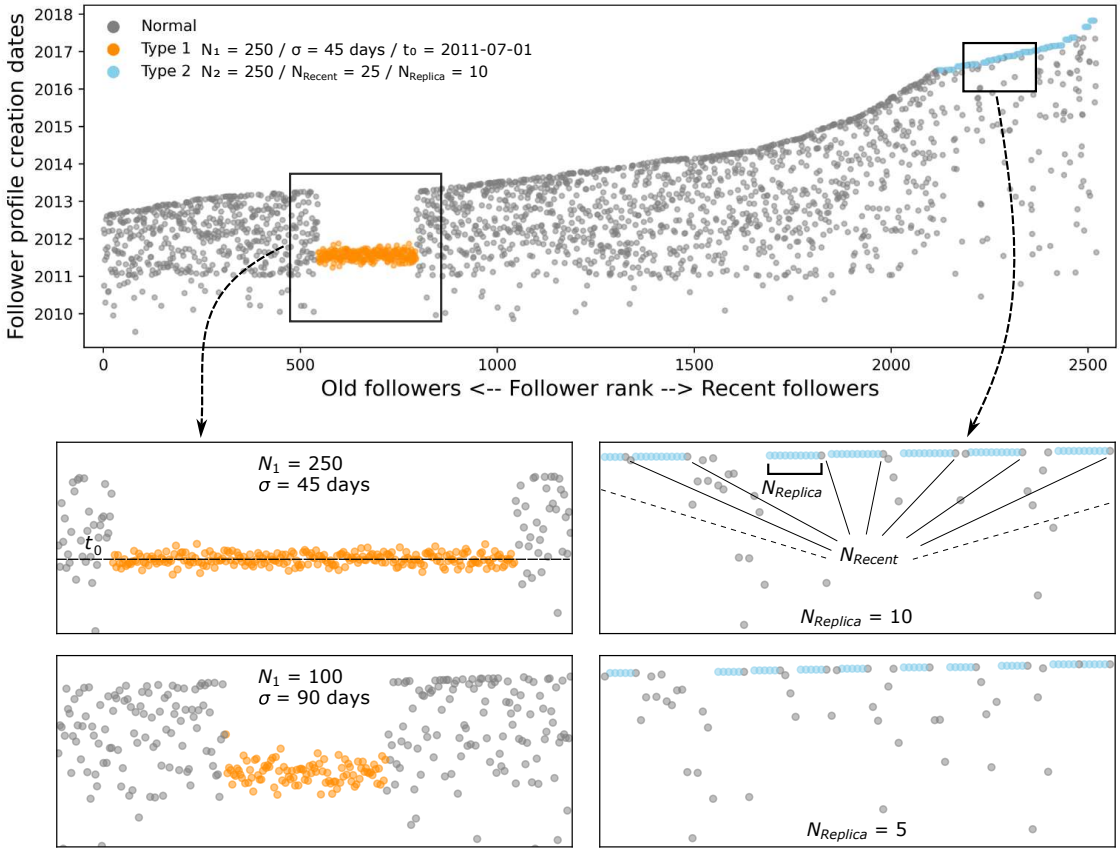

Figure 4.4 The two types of artificial anomalous followers and their corresponding parameters

Type 1	N_1	50, 100, 250, 500, 1000
	σ	10, 45, 90 days
Type 2	N_2	50, 100, 250, 500, 1000
	$N_{duplicate}$	5, 10

Table 4.1 Parameter values for synthetic follower generation.

4.4 Unsupervised Anomalous Follower Detection

We tested several existing unsupervised anomaly detection methods on the synthetic dataset. The performance of these methods on the synthetic dataset indicates how well these methods are applicable for detecting anomalous followers in real datasets. While we could have also used supervised learning models on the synthetic dataset, they may not be able to generalize well to real datasets since our synthetic dataset is a simplification of the real anomalous followers. The unsupervised anomaly detection methods that we tested are Isolation forest Liu, Ting & Zhou (2008), Local Outlier Factor (LOF) Breunig, Kriegel, Ng & Sander (2000), Empirical-Cumulative-distribution-based Outlier Detection (ECOD) Li, Zhao, Hu, Botta, Ionescu & Chen (2022), and Gen2Out Lee, Shekhar, Faloutsos, Hutson & Iasemidis (2021). For these methods, we generated features from the follower map that can help identify the anomalous followers (Section 4.4.1). Additionally, we presented a novel approach, Sliding Histogram (SH), that is specific to the task of anomalous follower detection.

4.4.1 Feature Engineering

As shown earlier, the anomalous followers we aim to detect can be visually identified on the follower map. They correspond to groups of consecutive followers of a user that have an abnormal distribution of profile creation dates. Thus, in order to apply anomaly detection algorithm to detect these followers, we created a set of features that can describe the local distribution around a follower in the follower map. The features used to detect anomalous followers using anomaly detection algorithms are described in Table 4.2. Fig. 4.5 demonstrates the lower and upper bounds of a

follower map, in addition to the window used to compute features that are based on the neighbors of a follower.

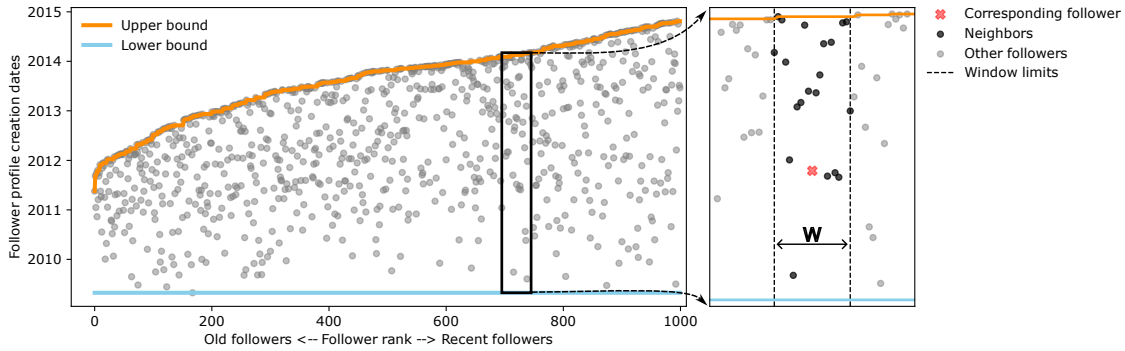

Figure 4.5 Profile creation date lower and upper bounds, and centered window around the corresponding follower for which features are computed

Feature	Description
Avg. neighbor creation date	The average profile creation date of the neighbors in a centered window of width W, weighted by the rank difference between each neighbor and the corresponding follower
Neighbor creation date range	Difference between the 90th and 10th percentiles of the creation dates of the neighbors in a centered window of width W
Avg. distance to neighbors	Average distance to the neighbors in a centered window of width W, measured in terms of creation dates and weighted by the rank difference between each neighbor and the corresponding follower
Creation date boundary range	The difference between the lower and upper bounds of profile creation dates at the rank of the corresponding follower
Distance to upper bound	Difference between the profile creation date upper bound and the profile creation date of the corresponding follower
Relative rank	Rank of the corresponding follower divided by the total number of followers

Table 4.2 Unsupervised anomaly detection feature definitions

4.4.2 Isolation Forest

In the isolation forest algorithm Liu et al. (2008), a forest of decision trees with random splits is grown, and higher anomaly scores are given to points that have a shorter average path in the forest. The path length is the number of splits from the root node required to isolate a data point in a leaf node. This definition of anomaly score is based on the fact that anomalies, by definition, are "few and different". Therefore, by randomly splitting nodes in a decision tree, we expect anomalies to be

isolated earlier than normal points since they reside in sparser areas of the feature space. Isolation forest trees are created using sub-samples of the dataset to avoid two common problems in anomaly detection: swamping and masking. The isolation forest algorithm requires two main hyperparameters: number of trees in the forest and sub-sample size. In our experiment, we use 200 trees and a sub-sample size of 256, which is the size recommended by the authors.

4.4.3 Local Outlier Factor

The Local Outlier Factor algorithm (LOF) Breunig et al. (2000), is designed to detect local outliers, i.e., points that lie in areas with less density than that of the nearest cluster of points. A point is assigned a high anomaly score if the average distance between this point and its nearest neighbors is greater than the average distance between its nearest neighbors and their nearest neighbors. The main hyperparameter in this algorithm is the number of nearest neighbors to be considered (*MinPts*). Since we are dealing with groups of anomalous followers, we expect them to be clustered together in the feature space. Thus, *MinPts* should be set to a value greater than the number of anomalies in a group of anomalous followers. Otherwise, this cluster of anomalies would be assigned low anomaly scores since all the nearest neighbors would be inside the same cluster. However, the fact that we do not have prior information about the number of anomalies that we expect to see in one group makes it hard to choose the value of *MinPts*. In our experiment, we set *MinPts* to 3% of the total number of followers of each user. Although users may have an anomaly ratio greater than 3% in their followers, larger values of *MinPts* result in prohibitive run times and memory usage for users with a large number of followers.

4.4.4 ECOD

The Empirical-Cumulative-distribution-based Outlier Detection (ECOD) method assigns high outlier scores to data points that have a low tail probability under the joint cumulative distribution function (CDF) of the data Li et al. (2022). The joint CDF is estimated by assuming that the dimensions (features) of the data are independent. Thus, the product of the univariate empirical CDFs (ECDF) of all dimensions is used as an estimate of the joint CDF. Data points that have extreme

feature values, based on the distribution of the corresponding feature, receive high outlier scores. This method does not require any hyperparameter tuning and is computationally efficient. However, due to the independence assumption, the interactions between features are not considered in this method.

4.4.5 Gen2Out

The Gen2Out method relies on the same concept of the IF method, i.e., an anomalous point tends to have a shorter average path from the root node to its leaf node in a forest of random decision trees, referred to as *AtomTrees* in this study Lee et al. (2021). However, instead of growing full trees on subsets of the dataset, trees are grown to a predefined maximum depth using all of the data points. The path length of each data point (q) to its leaf node is then estimated using Eq. 4.1, where h_0 is the path length up to the final node that the data point q falls in, l_{busy} is the number of points in that node, and $H(l_{busy})$ the estimated depth of an *AtomTree* grown using l_{busy} points.

$$h(q) = h_0 + H(l_{busy}) \tag{4.1}$$

The authors demonstrate that a linear relationship exists between the depth of the *AtomTree* and the logarithm of the count of data points used to construct the tree, regardless of the distribution of the data. Based on this observation, a number of *AtomTrees* are grown using several subsets of the data set to fit a linear function H that maps the logarithm of the count of points to the depth of a fully grown *AtomTree*. The anomaly score assigned to a point q is then computed using Eq. 4.2, where n is the number of points in the considered data set and $E[h(q)]$ is the average path length of point q in the forest.

$$s(q,n) = 2^{-\frac{E[h(q)]}{H(n)}} \tag{4.2}$$

4.4.6 Sliding Histogram

Our proposed approach specifically addresses anomalous groups defined in this study, i.e., dense groups of followers created in a tight time range. This is achieved by finding groups of followers that have a local distribution in the follower map that is significantly different from the overall distribution of the followers of the same user. The steps of this method are described as follows:

- A window with a predefined width (b) is slid along the rank axis of the follower map. The window stretches on the timestamp axis between the lower and upper bounds of the follower timestamps at that position (Fig. 4.6a).
- At each position, the window is divided into a predefined number of bins (N_{bins}) and the number of followers in each bin is computed (Fig. 4.6b). These histograms are shown as line plots in Fig. 4.6c.
- At each bin position, the median and inter-quartile range (IQR) of all histograms are computed.
- An anomaly score is assigned to each histogram bin using Eq. 4.3. Thus, each bin of followers is assigned a score that is the number of IQRs between the follower count in that bin and the median of follower counts in all bins at the same position.

$$A_{ij} = \frac{H_{ij} - M_j + 1}{IQR_j + 1} \tag{4.3}$$

  Where H_{ij} is the count of followers in the bin j of the window i, and M_j and IQR_j are the median and IQR of follower counts in the bin j across all windows, respectively.

- Since we are using a sliding window, each follower appears in more than one window. Thus, an anomaly score can be assigned to each individual follower f using a weighted average of all bin scores A_{ij} that include the follower f. The weight λ_{fi} (Eq. 4.4) takes its maximum value when the follower f is in the center of the bin and its minimum value when the follower f is at the edge of the bin. The anomaly score is then computed using Eq. 4.5

$$\lambda_{fi} = 1_{f \in W_i} \left(\frac{\frac{b}{2} - |R_f - C_i| + 1}{\sum_j \frac{b}{2} - |R_f - C_j| + 1} \right) \tag{4.4}$$

  Where b is the width of the sliding window, R_f is the rank of the follower f, and C is the center of the sliding window.

$$score_f = \sum_{j}^{N_{bins}} \sum_{i}^{N_{windows}} \lambda_{fi} A_{ij} 1_{f \in W_i} 1_{f \in B_{ij}} \tag{4.5}$$

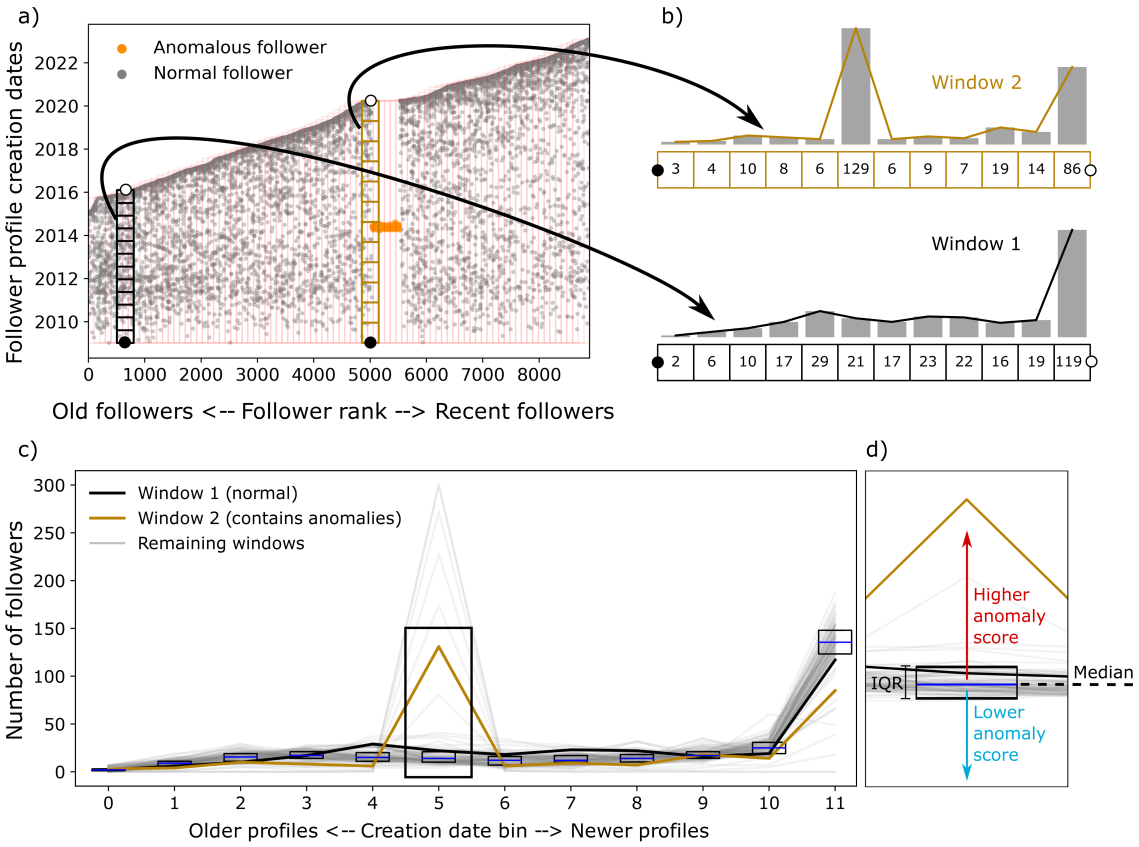

Figure 4.6 Illustration of the Sliding Histogram (a) A follower map with inserted synthetic irregular followers (orange) showing all sliding windows (light gray) with two of them highlighted in black and orange. (b) The histograms corresponding to the two highlighted windows in the follower map. Window 1 only includes normal followers and Window two includes anomalous followers. The numbers are the count of followers that fall within each bin. (c) All histograms plotted together as line plots, with the black and orange lines corresponding to the black and orange windows above. (d) A zoom in on bin No. 5 showing the median and interquartile range (IQR) of all histograms at this bin.

4.5 Follow-Time Estimation

Twitter does not provide the timestamps corresponding to when a user followed a certain account on Twitter. However, Twitter provides an ordered list of the followers of each account. Therefore, for a Twitter account, we know the order in which all of the account's followers have followed this account. Using this information, along with the profile creation dates of all followers, it is possible to estimate the follow times as shown in Meeder, Karrer, Sayedi, Ravi, Borgs & Chayes (2011). Given an account on Twitter with N followers, let C_i and F_i be the profile creation date and follow time of the i^{th} follower, respectively, where $i \in (1, N)$. We know that the follow time of a follower of an account is after the creation date of all previous followers of the same account, i.e., $F_i \geq max(C_1, C_2, ..., C_i)$. Thus, the value $max(C_1, C_2, ..., C_i)$ is a lower bound for the follow time of the i^{th} follower. Meeder et al. show that this lower bound can be used as an estimator for follow time with reasonable accuracy when the followed account has a high follow rate Meeder et al. (2011). We slightly modified this estimator by interpolating the follow time of each follower between its lower bound, i.e., $max(C_1, C_2, ..., C_i)$, and the next unique lower bound of subsequent followers. We evaluated this follow time estimation method on the Dribbble dataset which includes ground-truth values for follow times. As seen in Fig. 4.7, the average follow time estimation error across the followers of an account drops below one day when the account has more than 10,000 followers. Since we are dealing in this study with popular accounts (Fig. 3.1), the follow time estimation error will be well below one day in our data. We used the estimated follow times to conduct analyses on the coordinated behavior of anomalous followers.

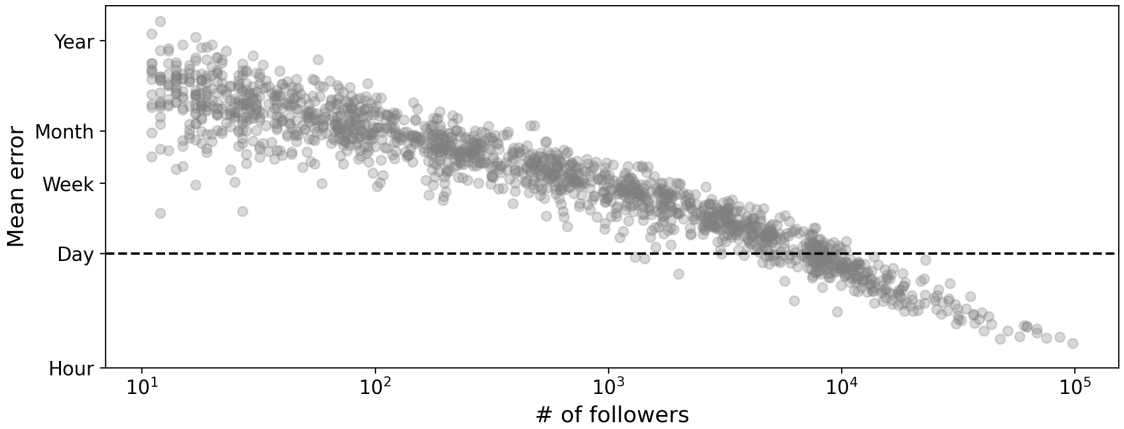

Figure 4.7 Mean follow time estimation error. Each point represents the error between the estimated follow time and the ground truth averaged across all followers of one Dribbble user. The mean error is less than one day for users with more than 10,000 followers.

5. Results

5.1 Results on synthetic data

The aim of the experiment on synthetic data is to identify the best performing detection method for detecting anomalous followers before applying it on real Twitter data. We experimented with the detection methods using 3 window sizes (51, 101, 201). The window size is a hyper parameter of the SH detection method, and it corresponds to the number of neighbors considered when creating features for the remaining methods. Table 5.1 shows the area under the ROC curve (AUC), average precision (AP), and precision when considering the highest 50 anomaly scores (precision@50) of the detection methods on the synthetic data generated from the Dribbble dataset. Note that the presented scores are averaged across all the synthetic cases ($55 \times 2,834$ synthetic cases), hence the high standard deviation values. Figures 5.1-5.2 provide a more detailed view of the performance of the ECOD and SH methods, respectively, on the synthetic dataset by showing the performance metrics against the ratio of anomalous followers in each synthetic case. It can be seen that our suggested method (SH) outperforms the feature-based methods regardless of the window size.

Table 5.1 Results on Dribbble dataset. Area under ROC curve, average precision, and precision at 50 mean (std) values for all methods using different window sizes.

Window	Method	AUC	AP	P@50
W51	ECOD	0.71 (0.22)	0.31 (0.13)	0.26 (0.21)
	Gen2Out	0.62 (0.31)	0.26 (0.10)	0.15 (0.18)
	IsolationForest	0.61 (0.30)	0.24 (0.11)	0.09 (0.17)
	LocalOutlierFactor	0.54 (0.20)	0.28 (0.18)	0.49 (0.31)
	SlidingHistogram	**0.86 (0.15)**	**0.69 (0.23)**	**0.72 (0.39)**
W101	ECOD	0.70 (0.21)	0.29 (0.13)	0.21 (0.19)
	Gen2Out	0.63 (0.30)	0.25 (0.11)	0.12 (0.18)
	IsolationForest	0.62 (0.28)	0.23 (0.11)	0.07 (0.16)
	LocalOutlierFactor	0.51 (0.19)	0.25 (0.18)	0.46 (0.37)
	SlidingHistogram	**0.87 (0.15)**	**0.71 (0.23)**	**0.72 (0.39)**
W201	ECOD	0.66 (0.21)	0.26 (0.13)	0.16 (0.17)
	Gen2Out	0.59 (0.27)	0.22 (0.12)	0.05 (0.11)
	IsolationForest	0.58 (0.26)	0.20 (0.12)	0.02 (0.09)
	LocalOutlierFactor	0.45 (0.17)	0.21 (0.16)	0.37 (0.38)
	SlidingHistogram	**0.87 (0.15)**	**0.69 (0.24)**	**0.70 (0.40)**

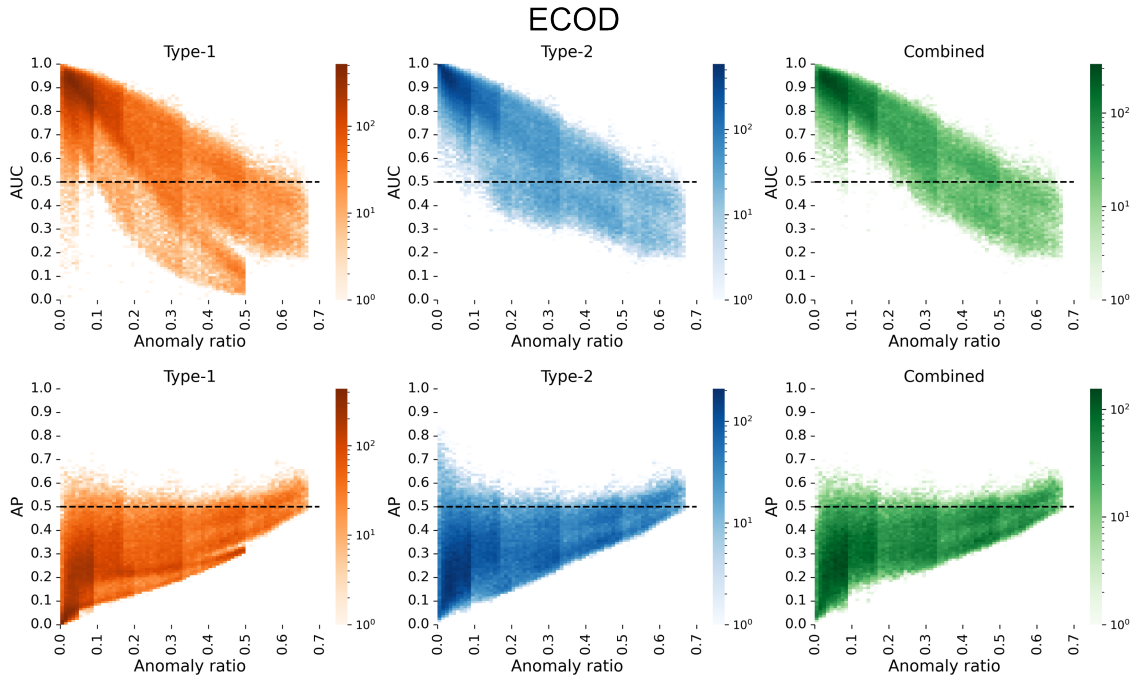

Figure 5.1 Heatmap of AUC (top) and AP (bottom) of the ECOD method for all the synthetic cases plotted by the ratio of anomalous followers to the total number of followers in each case. The plotted values correspond to the window size 101.

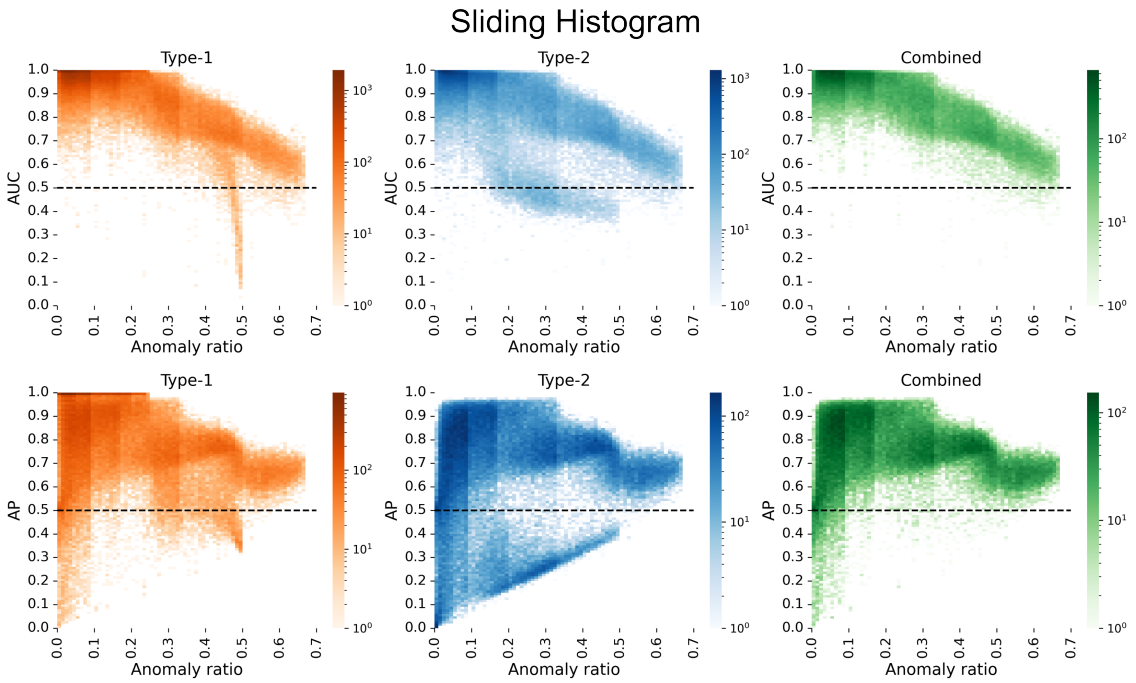

Figure 5.2 Heatmap of AUC (top) and AP (bottom) of the SH method for all the synthetic cases plotted by the ratio of anomalous followers to the total number of followers in each case. The plotted values correspond to the window size 101.

5.2 Results on real data

To verify the ability of our suggested method, we applied it on 1,318 accounts of Turkish politicians and media outlets from the #Secim2023 dataset **?**. We applied the SH method on these accounts using a window size of 200. The results of this analysis are presented in three parts: (i) Retrieving user accounts that have anomalous followers (ii) Identifying individual anomalous follower accounts (iii) Exploring the coordinated behavior of the detected anomalous followers.

5.2.1 Retrieving users with anomalous followers

In order to detect the users that have anomalous follow patterns among their followers, we first looked at the 9 Twitter accounts with the highest average anomaly score across all their followers (Fig. 5.3). The follower maps are shown here as heat maps

instead of scatter plots since these users have large numbers of followers. Irregular follow patterns can be observed in all of the follower maps of these users. Since the average anomaly score across all followers is generally lower for popular accounts, we can alternatively look at the average anomaly score of the highest N anomaly scores of a user's followers. Fig. 5.4 shows the anomalous followers of four popular Twitter accounts from our dataset. The anomalous following patterns in popular accounts cannot be visually observed on the follower map without zooming in. Figure 5.5 shows additional cases of obvious anomalous following patterns. It is important to notice that our approach was able to identify anomalous following patterns that are different from the synthetically created patterns that we experimented on. This property of our approach is attributed to its reliance on finding out-of-distribution following patterns rather than relying on predefined features specific to a certain pattern.

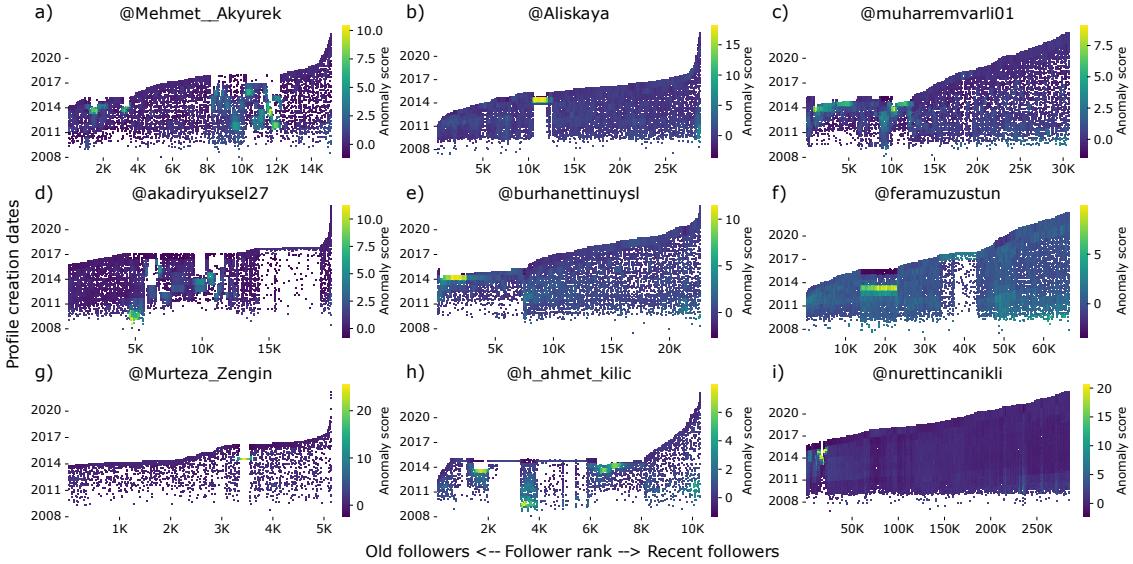

Figure 5.3 Retrieving users with anomalous followers Follower maps of the 9 Twitter accounts with the highest average anomaly score across all of their followers. The colors represent the average anomaly scores of all followers that fall in each bin (cell) of the heat map.

5.2.2 Identifying individual anomalous accounts

In this part, we examine the individual accounts in the detected groups of anomalous followers First, we looked at these accounts' bot scores as computed by Botometer-Lite Yang, Varol, Hui & Menczer (2020). The BotometerLite only uses features that can be extracted from the account information, making it applicable to our

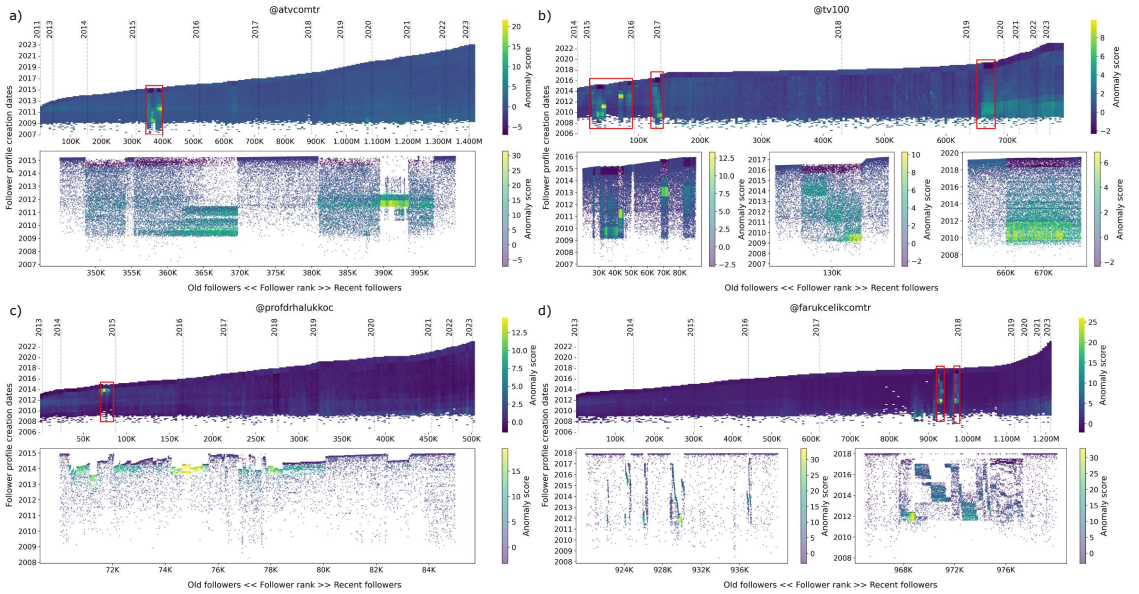

Figure 5.4 Follower maps of 4 popular Twitter accounts (>500k followers) with anomalous followers. The sub figures under each follower map are a zoom-in on the parts marked by a red box on the main follower map.

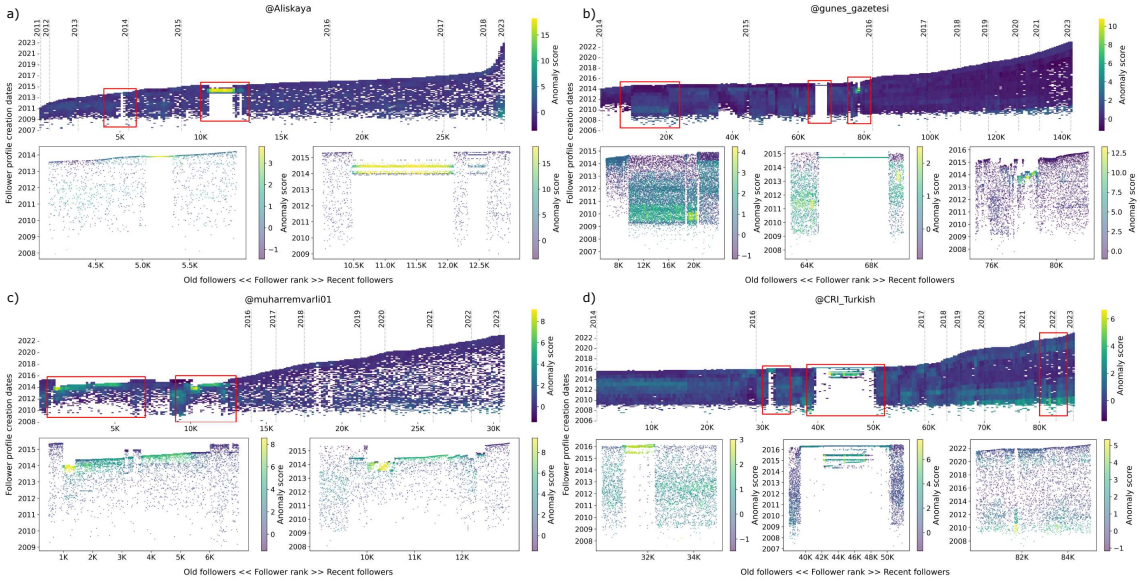

Figure 5.5 Follower maps of 4 Twitter accounts with obvious anomalous followers. The sub figures under each follower map are a zoom-in on the parts marked by a red box on the main follower map.

dataset. We refer to the scores computed by the BotometerLite as *bot scores*. Fig. 5.6 shows two cases, (A) anomalous followers having high bot scores (B) anomalous followers having low bot scores. To validate that the anomalous followers in the second case are indeed suspicious accounts, we manually observe a sample of these accounts. Appendix A presents samples of Twitter profiles of irregular followers of three accounts from our datasets, including the two accounts shown in Fig. 5.6, along with snapshots of these anomalous profile webpages on the Wayback Machine (Internet Archive). We observed that many of these accounts share the same tweets and share many of their friends. Additionally, the usernames of these accounts are in many cases meaningless combinations of letters. Fig. 5.6d and Fig. 5.6h show the distribution of the friend, follower, and status counts of the anomalous followers compared to that of all the followers of the same account. In both cases A and B, the anomalous accounts tend to have a lower number of followers. In case A, the anomalous followers have a low number of shared posts, indicating that they are mainly aimed at increasing the follower counts. On the other hand, the anomalous followers in case B share a lot of posts, indicating that they are used to spread information. These results show that our approach can capture bots that act in coordination, even though their bot scores as computed by other methods may not necessarily be high.

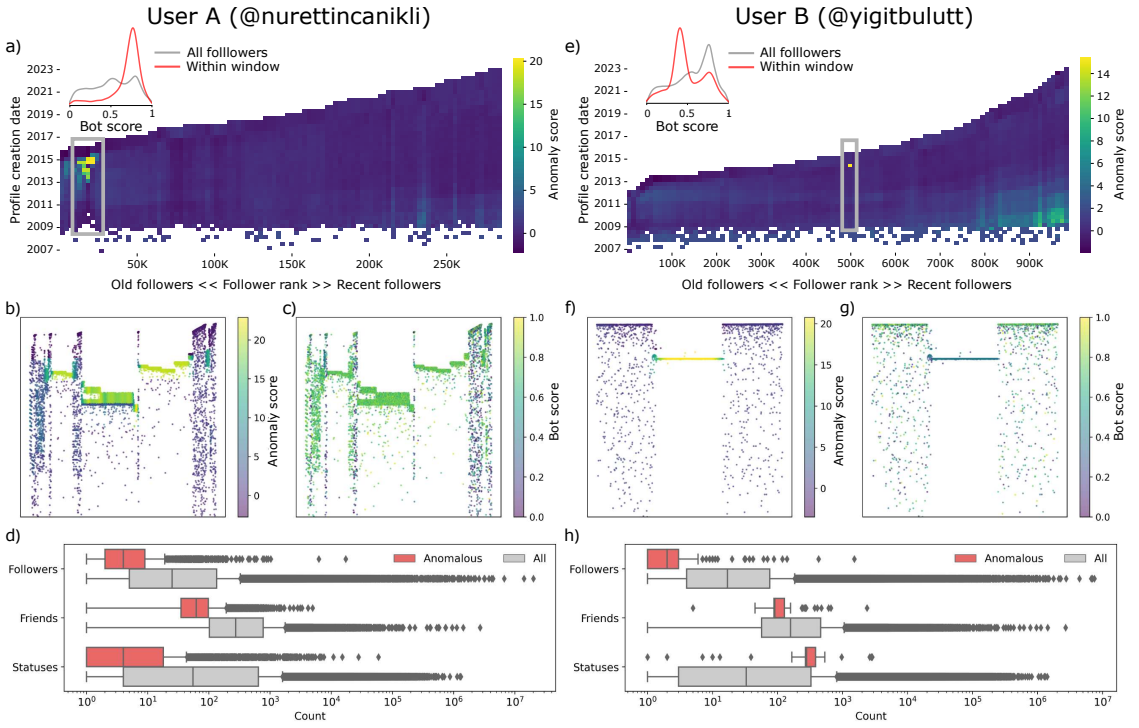

Figure 5.6 Detailed analysis of anomalous followers. User A: Anomalous followers have high bot scores. User B: Anomalous followers have low bot scores. Anomalous regions are zoomed in for User A (b,c) and User B (f,g). Profile statistics for regular and all followers are also compared for these users in subplots (d) and (e).

5.2.3 Exploring anomalous follower group behavior

We explored the following patterns of the detected groups of anomalous followers and studied when they follow other users in our Twitter dataset. Are they always showing suspicious following patterns for other politicians, or is it specific to the particular user that we made the observation for? Firstly, we looked for accounts in our dataset that are followed by at least 30% of the suspicious followers of users A and B (Fig.5.6). We found 0 accounts followed by the anomalous followers of user A and 12 accounts followed by the anomalous followers of user B. Since the anomalous accounts following user A do not follow any other users from our dataset, we resumed our analysis for user B only. We estimated the dates that the anomalous followers followed each of the 13 Twitter accounts using the method suggested in Meeder et al. (2011). Fig.5.7(a) and Fig.5.7(b) show the following times and anomaly scores, respectively, of the anomalous followers (red) and the followers shared across the 13 users (gray) for comparison. The anomalous followers follow each user almost simultaneously, which demonstrates that they are automated accounts that work in coordination. Furthermore, the anomalous followers followed all of the 13 users between the years 2014 and 2016. Finally, our approach correctly assigned high anomaly scores to the anomalous followers in most cases (Fig.5.7(b)).

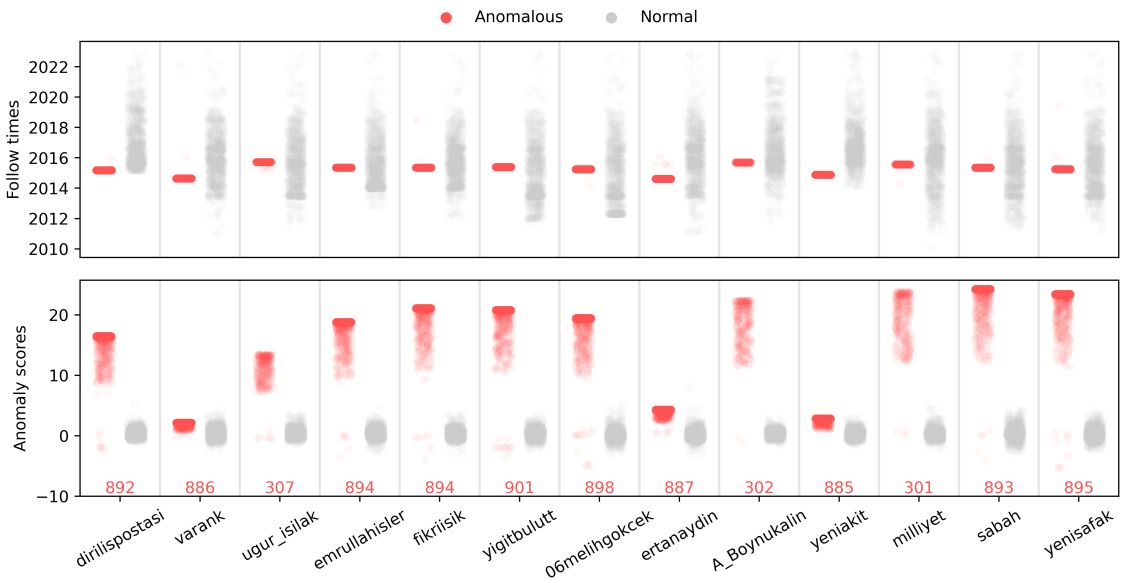

Figure 5.7 Coordinated behavior of anomalous followers. Follow times (top) and anomaly scores (bottom) of the shared anomalous followers (red) and the shared non-anomalous followers (gray) across 13 users that are followed by the same batch of anomalous followers shown in Fig.5.6(f).

We expanded the analysis of the group behavior of anomalous followers to uncover other groups of accounts that share the same suspicious followers. For this pur-

pose, we created a similarity network based on the shared anomalous followers. The similarity between each pair of accounts is the cosine similarity between the two anomaly score vectors of the followers shared across the pair of users. Since our method assigns anomaly scores based on the follower map, a follower that follows accounts U1 and U2 will have two different anomaly scores computed for U1 and U2. Thus, a pair of accounts that share followers who were assigned high anomaly scores in both follower maps will have a high similarity. The Louvain community detection algorithm was then used to detect the communities in the networkBlondel, Guillaume, Lambiotte & Lefebvre (2008). Fig.5.8 shows the two communities with the highest pairwise average anomaly scores across all edges in the community. For each community, we show the follower maps of a user pair corresponding to one of the edges in the community. The follower maps are colored by the ratio of shared followers between the pair of users in each bin. This allows us to capture concentrations of shared followers in both users' maps, which appear as reddish regions in the follower map. We observe that the concentrated regions of shared followers exhibit anomalous following patterns in both follower maps. This finding supports our hypothesis that anomalous followers work in coordination. More details about this network analysis and other samples of anomalous follower groups appearing in different users' follower maps are presented in Appendix A.

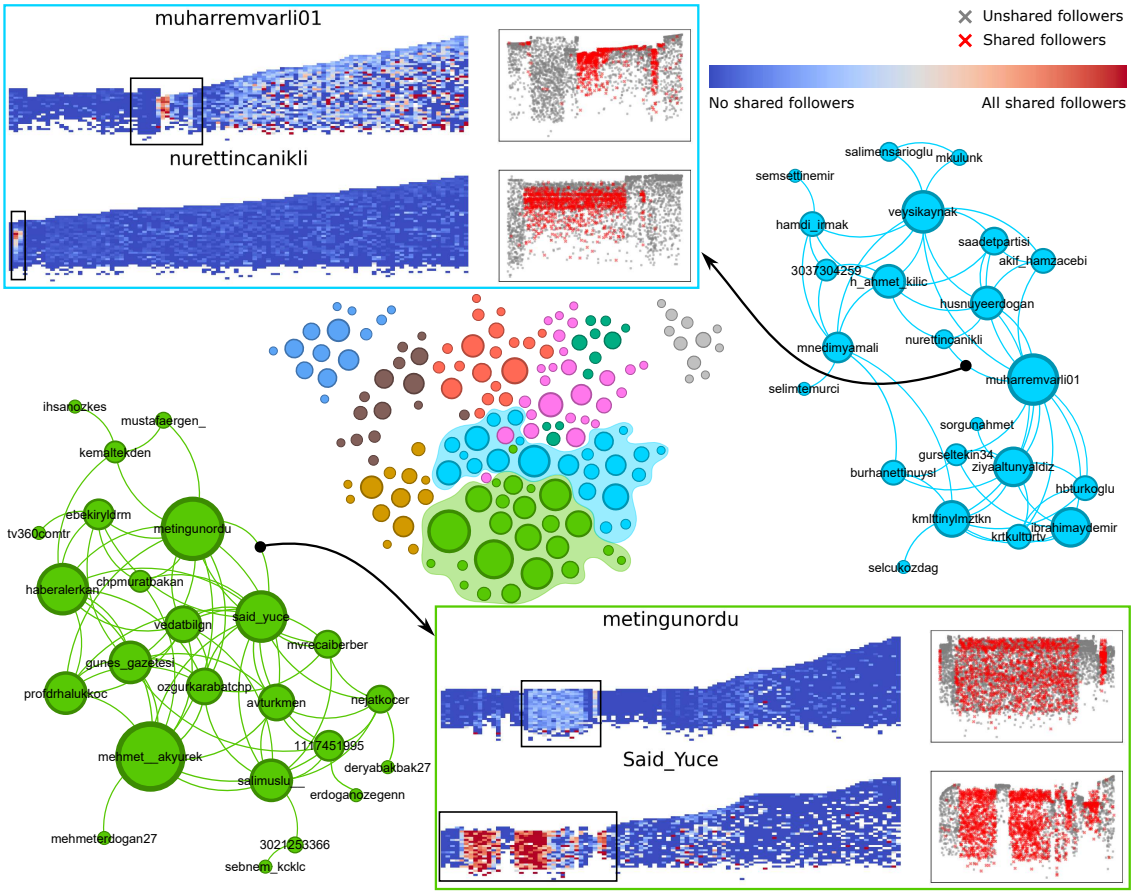

Figure 5.8 Similarity network based on the shared anomalous followers. The full network is shown in the middle of the figure, where nodes are colored based on communities and sized based on their degrees. The two communities with the highest pairwise average anomaly scores are highlighted and shown in detail along with the follower maps of one edge in each community.

6. Discussion

6.1 Anomalous follower detection

The term "anomalous followers" can refer to any social media follower that exhibits a behavior that is considered abnormal for ordinary social media users. In this study, we focused on a collective abnormal behavior to define the anomalous followers that we aim to detect. This collective abnormal behavior can be identified by observing the temporal patterns of users following a certain account in relation to their profile creation dates, which we represent using the follower map. This type of abnormal behavior was noticed in previous works Bellutta & Carley (2023); Confessore et al. (2018); Varol & Uluturk (2020), but has not been addressed specifically by any detection methods. The detection method we introduced in this study specifically targets followers that exhibit this type of abnormal behavior. Our method differs from other detection methods in two main aspects: (i) It is applied on all the followers of a social media account at once (ii) The anomaly score it assigns to each follower is unique to this follower following this specific user, i.e., a follower following 10 users will have 10 different anomaly scores each corresponding to one of the users. These characteristics of our proposed method make it suitable for: (i) identifying users that have anomalous followers (ii) investigating the coordinated behavior of anomalous followers across multiple accounts (iii) identifying anomalous followers with high confidence. The latter point is attributed to the fact that our method assigns multiple scores to each follower. As for any detection method, a follower may be falsely assigned a high anomaly/bot score. However, by looking at the follower's anomaly score across several users, we can increase our confidence about the abnormality of this follower. As for investigating coordinated behavior, the definition of anomalous followers that we use in itself implies coordinated behavior. The study of this coordinated behavior can be extended by observing the behavior of an anoma-

lous group of followers across several accounts. Furthermore, coupling the anomaly scores with follow-time estimates opens the door for investigating misinformation campaigns Bessi & Ferrara (2016); Hristakieva et al. (2022); Wu et al. (2019) and astroturfing campaigns Keller et al. (2020); Schoch et al. (2022) at specific dates in the past.

6.2 The case of Turkish political Twitter

Applying our method on Twitter users from the Turkish political sphere showed that numerous politicians are followed by anomalous follower accounts that are still active on Twitter. Furthermore, by creating a network between Twitter accounts based on their shared anomalous followers, we identified groups of anomalous followers that are active across several accounts. Additionally, we observed that different groups of politician/media accounts share different groups of anomalous followers. We do not make any conclusions about the intention of the coordinated anomalous followers. However, we hypothesize three possible scenarios: (i) The anomalous followers were purchased by the user to manipulate popularity metrics (ii) The anomalous followers chose to follow specific user to portray a certain image of themselves (iii) The anomalous followers targeted the user to initiate a smear campaign against the user in the benefit of the user's opponents. Future studies that look into the time of these campaigns and the political leanings of the users may provide more insights on the purpose of these anomalous followers.

6.3 Limitations

The detection method we suggest in this study relies on detecting groups of followers that have a temporal follow-pattern in the follower map that deviates from the general pattern in the same follower map. Therefore, this method fails by design when the user has more anomalous followers than normal ones. Fig. 6.1 shows the follower map and anomaly scores of a number of users that have a high ratio of anomalous followers. In such cases, high anomaly scores are assigned wrongly to normal followers. Nevertheless, these users will still have a high average anomaly

score amongst their followers, albeit wrongly assigned. This makes it possible to identify them as users with anomalous followers among a pool of users.

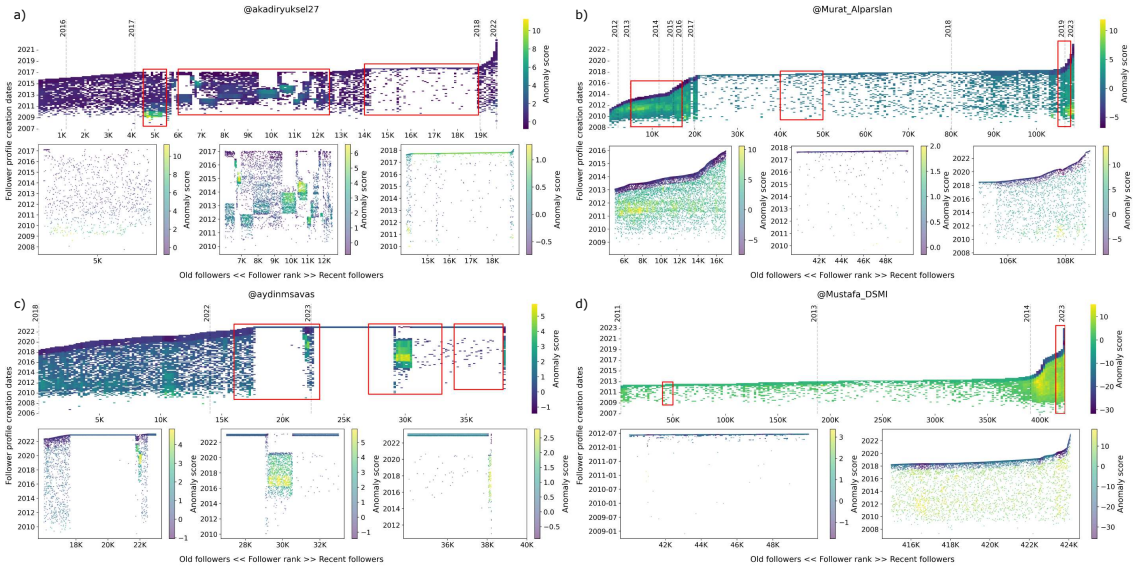

Figure 6.1 Users that have a high ratio of anomalous followers.

7. Conclusion

In this study, we introduced a novel unsupervised method for the detection of anomalous followers on social media platforms. Our approach can be applied to all the followers of a social media user using only an ordered list of these followers and their profile creation dates. Furthermore, we applied our detection method on a real dataset from Twitter comprising 1,318 accounts of Turkish politicians and media outlets. We showed that, given a large pool of accounts, our introduced method is capable of identifying accounts that are followed by anomalous followers. Moreover, our findings indicate that anomalous followers are prevalent on Turkish political Twitter. Additionally, our analyses identified a clear coordinated behavior of these anomalous followers across multiple accounts. While our study only focused on the follower-ship behavior of these anomalous followers, manual observation of their accounts showed that they tend to share similar content. Future studies may incorporate data about the activities of these accounts on Twitter to further investigate their coordinated behavior and their possible role in misinformation campaigns. Our detection method along with an implementation of the follow-time estimation algorithm are openly available in Python code: github.com/ViralLab/FollowerAnalyzer.

BIBLIOGRAPHY

Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, (pp. 839–851).

Alkulaib, L., Zhang, L., Sun, Y., & Lu, C.-T. (2022). Twitter bot identification: An anomaly detection approach. In *2022 IEEE International Conference on Big Data (Big Data)*, (pp. 3577–3585). IEEE.

Anstead, N. & O'Loughlin, B. (2015). Social media analysis and public opinion: The 2010 uk general election. *Journal of computer-mediated communication*, *20*(2), 204–220.

Auxier, B. & Anderson, M. (2021). Social media use in 2021. *Pew Research Center*, *1*(1), 1–4.

Bellutta, D. & Carley, K. M. (2023). Investigating coordinated account creation using burst detection and network analysis. *Journal of big Data*, *10*(1), 1–17.

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, *10*(2), e0118093.

Bessi, A. & Ferrara, E. (2016). Social bots distort the 2016 us presidential election online discussion. *First monday*, *21*(11-7).

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, *57*(2), 556–578.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, (pp. 93–104).

Bruno, M., Lambiotte, R., & Saracco, F. (2022). Brexit and bots: characterizing the behaviour of automated accounts on twitter during the uk election. *EPJ Data Science*, *11*(1), 17.

Chavoshi, N., Hamooni, H., & Mueen, A. (2016). Debot: Twitter bot detection via warped correlation. In *Icdm*, volume 18, (pp. 28–65).

Cinelli, M., Etta, G., Avalle, M., Quattrociocchi, A., Di Marco, N., Valensise, C., Galeazzi, A., & Quattrociocchi, W. (2022). Conspiracy theories and social media platforms. *Current Opinion in Psychology*, *47*, 101407.

Confessore, N., Dance, G. J., Harris, R., & Hansen, M. (2018). The Follower Factory. `https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html`. [Online; accessed 04-December-2018].

Cresci, S. (2020). A decade of social bot detection. *Communications of the ACM*, *63*(10), 72–83.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2015). Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, *80*, 56–71.

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017a). The

paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion*, (pp. 963–972).

Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017b). Social fingerprinting: detection of spambot groups through dna-inspired behavioral modeling. *IEEE Transactions on Dependable and Secure Computing*, *15*(4), 561–576.

Deb, A., Luceri, L., Badaway, A., & Ferrara, E. (2019). Perils and challenges of social media and election manipulation analysis: The 2018 us midterms. In *Companion proceedings of the 2019 world wide web conference*, (pp. 237–247).

DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PloS one*, *8*(11), e79449.

Ding, J. & Chen, Z. (2023). How to find social robots exactly? In *Proceedings of the 2023 6th International Conference on Software Engineering and Information Management*, (pp. 12–18).

Echeverrï£¡ a, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Stringhini, G., & Zhou, S. (2018). Lobo: Evaluation of generalization deficiencies in twitter bot classifiers. In *Proceedings of the 34th annual computer security applications conference*, (pp. 137–146).

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, propaganda, and disinformation: Online media and the 2016 us presidential election. *Berkman Klein Center Research Publication*, *6*.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96–104.

Ferrara, E., Wang, W.-Q., Varol, O., Flammini, A., & Galstyan, A. (2016). Predicting online extremism, content adopters, and interaction reciprocity. In *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part II 8*, (pp. 22–39). Springer.

Himelein-Wachowiak, M., Giorgi, S., Devoto, A., Rahman, M., Ungar, L., Schwartz, H. A., Epstein, D. H., Leggio, L., & Curtis, B. (2021). Bots and misinformation spread on social media: Implications for covid-19. *Journal of medical Internet research*, *23*(5), e26933.

Hristakieva, K., Cresci, S., Da San Martino, G., Conti, M., & Nakov, P. (2022). The spread of propaganda by coordinated communities on social media. In *Proceedings of the 14th ACM Web Science Conference 2022*, (pp. 191–201).

Jia, J., Wang, B., & Gong, N. Z. (2017). Random walk based fake account detection in online social networks. In *2017 47th annual IEEE/IFIP international conference on dependable systems and networks (DSN)*, (pp. 273–284). IEEE.

Jungherr, A. (2014). Twitter in politics: a comprehensive literature review. *Available at SSRN 2865150*.

Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, *13*(1), 72–91.

Keller, F., Schoch, D., Stier, S., & Yang, J. (2017). How to manipulate social media: Analyzing political astroturfing using ground truth data from south korea. In *Proceedings of the international AAAI conference on web and social media*, volume 11, (pp. 564–567).

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twit-

ter: How to coordinate a disinformation campaign. *Political communication*, *37*(2), 256–280.

Lee, M.-C., Shekhar, S., Faloutsos, C., Hutson, T. N., & Iasemidis, L. (2021). Gen 2 out: Detecting and ranking generalized anomalies. In *2021 IEEE International Conference on Big Data (Big Data)*, (pp. 801–811). IEEE.

Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., & Chen, G. (2022). Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Transactions on Knowledge and Data Engineering*.

Liedke, J. & Matsa, K. E. (2022). Social media and news fact sheet. pew research center.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, (pp. 413–422). IEEE.

Liu, Y., Tan, Z., Wang, H., Feng, S., Zheng, Q., & Luo, M. (2023). Botmoe: Twitter bot detection with community-aware mixtures of modal-specific experts. *arXiv preprint arXiv:2304.06280*.

Magelinski, T., Ng, L., & Carley, K. (2022). A synchronized action framework for detection of coordination on social media. *Journal of Online Trust and Safety*, *1*(2).

Mannocci, L., Cresci, S., Monreale, A., Vakali, A., & Tesconi, M. (2022). Mulbot: Unsupervised bot detection based on multivariate time series. In *2022 IEEE International Conference on Big Data (Big Data)*, (pp. 1485–1494). IEEE.

Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., & Tesconi, M. (2019). Rtbust: Exploiting temporal patterns for botnet detection on twitter. In *Proceedings of the 10th ACM conference on web science*, (pp. 183–192).

Meeder, B., Karrer, B., Sayedi, A., Ravi, R., Borgs, C., & Chayes, J. (2011). We know who you followed last summer: inferring social link creation times in twitter. In *Proceedings of the 20th international conference on World wide web*, (pp. 517–526).

Mendoza, M., Tesconi, M., & Cresci, S. (2020). Bots in social and interaction networks: detection and impact estimation. *ACM Transactions on Information Systems (TOIS)*, *39*(1), 1–32.

Metaxas, P. T. & Mustafaraj, E. (2012). Social media and the elections. *Science*, *338*(6106), 472–473.

Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. *Journal of Cyber Policy*, *3*(1), 39–43.

Muhammed T, S. & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. *International journal of data science and analytics*, *13*(4), 271–285.

Najafi, A., Mugurtay, N., Zouzou, Y., Demirci, E., Demirkiran, S., Karadeniz, H. A., & Varol, O. (2024). First public dataset to study 2023 turkish general election. *Scientific Reports*, *14*(1), 8794.

Najafi, A. & Varol, O. (2024a). Turkishbertweet: Fast and reliable large language model for social media analysis. *Expert Systems with Applications*, 124737.

Najafi, A. & Varol, O. (2024b). Vrllab at hsd-2lang 2024: Turkish hate speech detection online with turkishbertweet. In *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, (pp. 185–189).

Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S., & Tesconi, M. (2021). Coordinated

behavior on social media in 2019 uk general election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, (pp. 443–454).

Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., & Menczer, F. (2021). Uncovering coordinated networks on social media: methods and case studies. In *Proceedings of the international AAAI conference on web and social media*, volume 15, (pp. 455–466).

Perrin, A. (2015). Social media usage. *Pew research center*, *125*, 52–68.

Pierri, F., Luceri, L., Jindal, N., & Ferrara, E. (2023). Propaganda and misinformation on facebook and twitter during the russian invasion of ukraine. In *Proceedings of the 15th ACM web science conference 2023*, (pp. 65–74).

Qudar, M. M. A. & Mago, V. (2020). Tweetbert: a pretrained language representation model for twitter text analysis. *arXiv preprint arXiv:2010.11091*.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Detecting and tracking political abuse in social media. In *Proceedings of the International AAAI Conference on Web and social media*, volume 5, (pp. 297–304).

Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2020). Detection of novel social bots by ensembles of specialized classifiers. In *Proceedings of the 29th ACM international conference on information & knowledge management*, (pp. 2725–2732).

Schoch, D., Keller, F. B., Stier, S., & Yang, J. (2022). Coordination patterns reveal online political astroturfing across the world. *Scientific reports*, *12*(1), 4572.

Seckin, O. C., Atalay, A., Otenen, E., Duygu, U., & Varol, O. (2024). Mechanisms driving online vaccine debate during the covid-19 pandemic. *Social Media+ Society*, *10*(1), 20563051241229657.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, *96*, 104.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications*, *9*(1), 1–9.

Sharma, K., Zhang, Y., Ferrara, E., & Liu, Y. (2021). Identifying coordinated accounts on social media through hidden influence and group behaviours. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, (pp. 1441–1451).

Singh, K., Lima, G., Cha, M., Cha, C., Kulshrestha, J., Ahn, Y.-Y., & Varol, O. (2022). Misinformation, believability, and vaccine acceptance over 40 countries: Takeaways from the initial phase of the covid-19 infodemic. *Plos one*, *17*(2), e0263381.

Takacs, R. & McCulloh, I. (2019). Dormant bots in social media: Twitter and the 2018 us senate election. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, (pp. 796–800).

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*.

Varol, O., Davis, C. A., Menczer, F., & Flammini, A. (2018). Feature engineering

for social bot detection. *Feature engineering for machine learning and data analytics*, *311*.

Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the international AAAI conference on web and social media*, volume 11, (pp. 280–289).

Varol, O. & Uluturk, I. (2020). Journalists on twitter: self-branding, audiences, and involvement of bots. *Journal of Computational Social Science*, *3*(1), 83–101.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social science & medicine*, *240*, 112552.

Weber, D. & Neumann, F. (2021). Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining*, *11*(1), 111.

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD explorations newsletter*, *21*(2), 80–90.

Yang, K.-C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., & Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1*(1), 48–61.

Yang, K.-C., Varol, O., Hui, P.-M., & Menczer, F. (2020). Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, (pp. 1096–1103).

Zhang, J., Carpenter, D., & Ko, M. (2013). Online astroturfing: A theoretical perspective.

Zouzou, Y. & Varol, O. (2023). Unsupervised detection of coordinated fake-follower campaigns on social media. *arXiv preprint arXiv:2310.20407*.

APPENDIX A

Anomalous profile samples

Figure A.1 Sample of anomalous followers of @nurettincanikli. The follower map of this user is shown in Fig. 5.7a

Figure A.2 Sample of anomalous followers of @yigitbulutt. The follower map of this user is shown in Fig. 5.7e

Figure A.3 Sample of anomalous followers of @matillakaya. The follower map of this user is shown in Fig. A.6

Username	Anomalous follower Wayback Machine link
@yigitbulutt	web.archive.org/web/https://twitter.com/bisetoveribo
	web.archive.org/web/https://twitter.com/hozuwocidob
	web.archive.org/web/https://twitter.com/lucemuhyzade
	web.archive.org/web/https://twitter.com/jyjehejuxok
@nurettincanikli	web.archive.org/web/https://twitter.com/786846f1e3ee48e
	web.archive.org/web/https://twitter.com/OuaHind
	web.archive.org/web/https://twitter.com/RogrioBellinca1
	web.archive.org/web/https://twitter.com/AbongJinky
@matillakaya	web.archive.org/web/https://twitter.com/dental654321
	web.archive.org/web/https://twitter.com/hacker_italy
	web.archive.org/web/https://twitter.com/Ezanaatt
	web.archive.org/web/https://twitter.com/FatihAk31652640

Table A.1 Internet Archive Wayback Machine links to the anomalous follower profiles presented in Fig. A.1-A.3.

Shared anomalous followers network

We create a user similarity network between the 1318 Twitter accounts in our dataset to observe the shared batches of anomalous followers between different users. The edge weight between each pair of accounts is the cosine similarity of the anomaly score vectors of the shared followers as computed from each of the follower maps of the pair of users. The six pairs of accounts corresponding to the highest six similarity scores are shown in Fig. A.4. The follower heat map colors represent the ratio of shared followers in each bin. Each pair of followers shown in Fig. A.4 share a group of followers that are concentrated in one area of the map, i.e., accounts that followed the user consecutively. The follow patterns of these batches of shared followers are anomalous as seen in the zoomed sub-figures. Fig. A.5 shows the user similarity network generated by filtering out all edge weights less than 0.75 and all edges with less than 100 shared followers. Nodes are sized by their degree and colored by their community membership. The Louvain community detection algorithm was used Blondel et al. (2008), which is based on modularity optimization. Fig. A.6 shows the community with the third highest pairwise average anomaly score (the first two visualized in the main text) and samples of the follower maps of connected user pairs. We observe that groups of anomalous followers that exhibit the same following pattern are observed in the followers of several accounts.

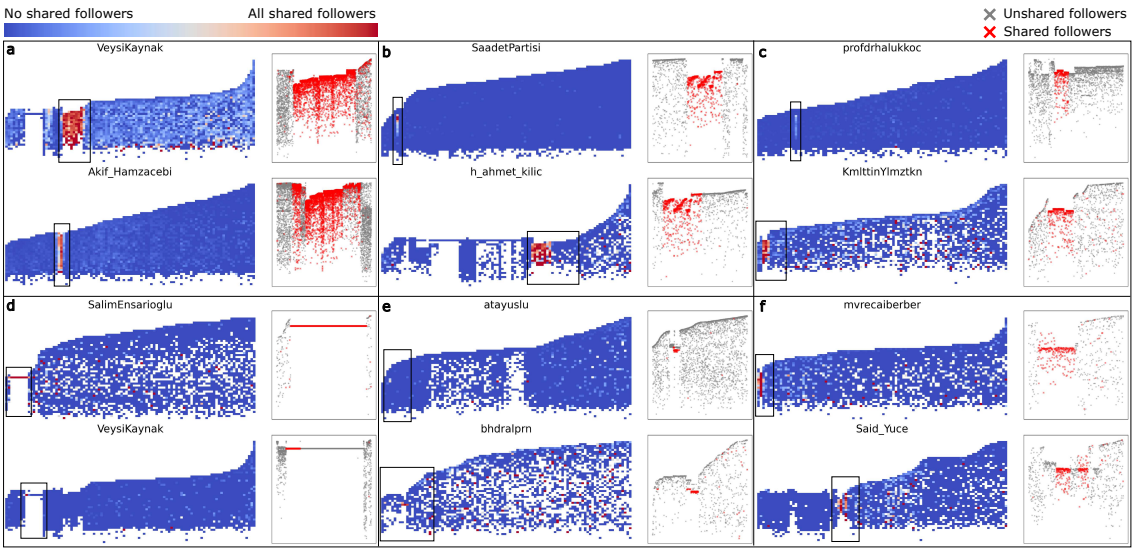

Figure A.4 Shared anomalous followers. Follower maps of the 6 user pairs corresponding to the highest similarity scores in our dataset.

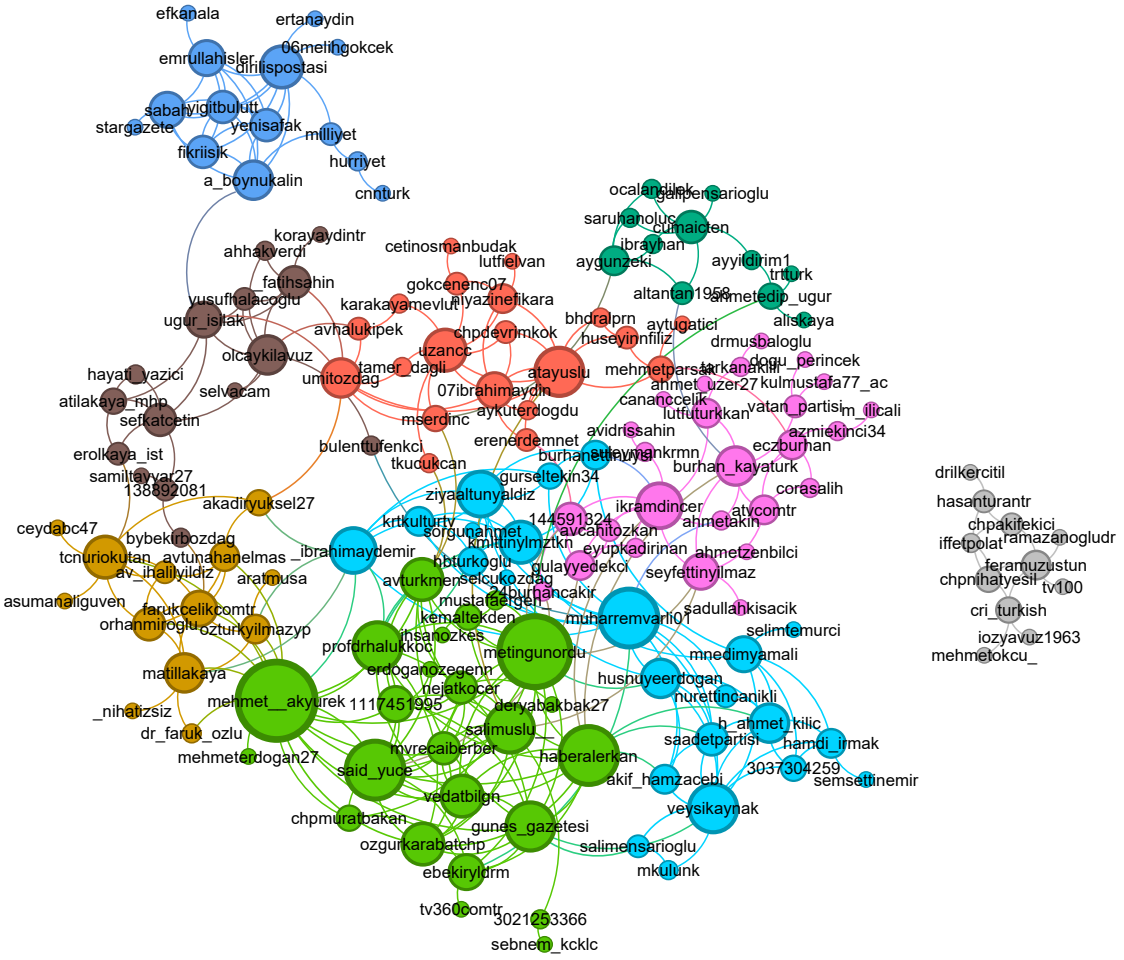

Figure A.5 Network of shared anomalous followers. Node colors represent community membership and node sizes are scaled by node degrees

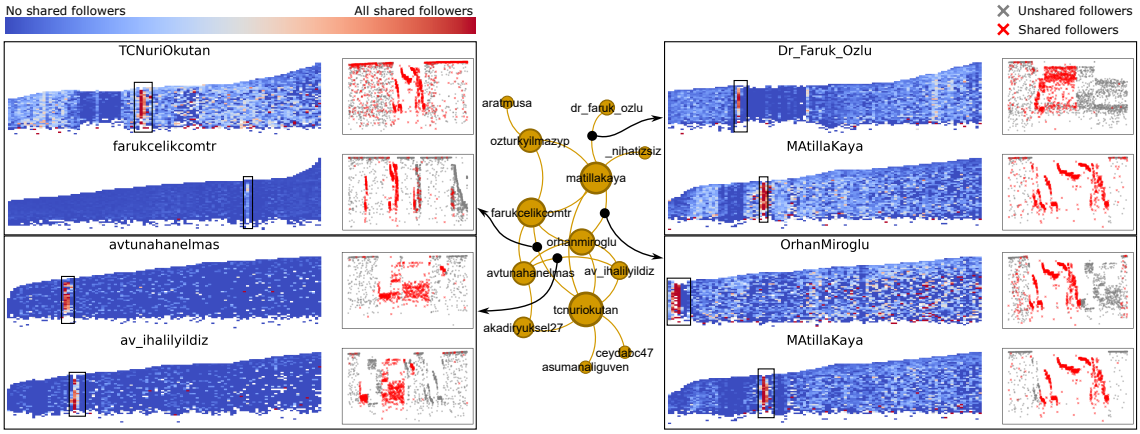

Figure A.6 Follower maps of user pairs corresponding to 4 edges in one of the communities of the shared anomalous followers network. The follow pattern of these anomalies are similar across several users.