# List of Figures


1.1 Directed link on Twitter . . . . . 3
1.2 Example of Fake Accounts . . . . . 4

2.1 Data Assembly . . . . . 8

3.1 Collusion Network Example . . . . . 13
3.2 Final Results . . . . . 18

4.1 Sample LDA Result . . . . . 23

5.1 Normal vs Malicious Accounts . . . . . 27
5.2 Correlation Matrix . . . . . 28
5.3 Twitter Signup . . . . . 30
5.4 Sum of mutually exclusive followers . . . . . 31
5.5 Top 21 Accounts . . . . . 35

6.1 Growth of Target Account's Followers . . . . . 38
6.2 Distribution of Zero Tweeters . . . . . 39

# List of Tables

2.1 **Data Acquisition for Collusion Networks** . . . . . 7
2.2 **Data Acquisition for Profile Analysis** . . . . . 7

3.1 **Symbols and their descriptions** . . . . . 12

4.1 **Results of Collusion Network** . . . . . 20

## 0.1 Acknowledgements

First and foremost, I would like to thank the individual who has mustered the courage to open up this thesis and intends to read it. I am well aware that endurance is seldom a tasteful way of grappling with boring ideas like these, so I do not expect you to read it till the end. I wouldn't, if I were you. The word endurance leads me to the thought that it must have been a similar experience for Dr.Fareed while working with me. I might have never added to his knowledge but I have definitely increased his patience level by continuously testing it. I truly acknowledge his contribution and guidance. I wish I knew better ways to thank him.

I also extend my gratitude to Dr. Jadoon for his guidance and support during the entire process. Along with him, I would like to thank Dr. Zartash for vividly enlightening me with the philosophy that smooth roads never make good drivers. However having a strict warden across the road might propel you to quit the idea of getting behind the wheel ever again. Having Dr. Zartash around, felt like a free scythe in my direction. Often feeling arrested by this fear, I found my inspiration of hard work in Dr Ijaz Naqvi. In our line of work, resurrection is impossible without a ritual of intellectual exorcism.

For all the good reasons, I would thank my parents for giving me the freedom to make all my choices in life. Now that I know how rare this happens in our culture, I am more inclined to cherish it for lifetime. It is indeed essential to create an open environment and belief system at home and let the off-springs make the most out of it. If it doesn't always work, it never completely fails. So a huge ThankYou to my amazing parents.

Finally, my warm regards to Javed Uncle. The finest intellectual I have ever seen. I often find myself in the safekeeping of his imperishable thoughts.

# Abstract

Twitter is one of the leading online social networks with an estimate of 328 million monthly active users. Among these users are various celebrities, politicians, world leaders and popular figures. With increasing popularity there is also a well known criticism regarding legitimacy and authenticity of activities going on this platform. The sign up features on Twitter facilitate a number of anomalous activities to take place. These activities include various black market promotion techniques such as fake followers, compromised accounts, cyborgs, collusion networks and malicious third party applications. This thesis demonstrates how such activities are carried out for reputation manipulation and promulgation of narratives across Twitter. We analyze engineered spread of information though lockstep Retweeting by bots and real users. Part of our study also involves exploring into the new phenomenon of "Daily Followers" which is an evolved version of acquiring fake followers. We also suggest ways in which malicious activities can be countered to reduce the effects of fake activities. Our major contribution is discovering collusion networks on Twitter which engage in a synchronized pattern of Retweeting and through these collusion networks we find out the associated beneficiary networks. From the text corpus scraped through these networks, we analyze popular themes being projected for normal users. We also build upon previous work of catching fake accounts and distinguishing network of fake accounts from real ones. In our research we demonstrate the modes of operations of underground markets which provide fake followers. We do nomenclature analysis to show a disguised fake accounts community that exists on Twitter's servers but never appears to the sight of a normal viewer at the front end.

# Contents


0.1 Acknowledgements . . . . . ix

**Abstract** **x**

**1 Introduction** **1**
1.1 Twitter as an OSN . . . . . 2
1.1.1 Links on Twitter . . . . . 3
1.2 Engineered Popularity on Twitter . . . . . 3
1.2.1 Organization . . . . . 4

**2 Background Research and Dataset** **5**
2.1 RELATED WORK . . . . . 6
2.1.1 Classical Approach . . . . . 6
2.2 Research on Fake Followers . . . . . 6
2.3 Acquiring Dataset . . . . . 7

**3 Collusion Networks** **9**
3.1 COLLUSION NETWORKS on TWITTER . . . . . 10
3.1.1 Pretext . . . . . 10
3.1.2 Previous Work . . . . . 11
3.1.3 Formulation . . . . . 11
3.2 EXPERIMENT on COLLUSION NETWORKS and RESULTS . . . 13
3.2.1 Use of *CopyCatch* . . . . . 14
3.2.2 Detecting Beneficiary Network . . . . . 14
3.2.3 Tools . . . . . 15
3.2.4 Interpretation . . . . . 15

**4 Monitoring Narrative** **19**
4.0.1 Monitoring Narrative . . . . . 20
4.0.2 LDA Example . . . . . 21
4.0.3 Sample LDA Result . . . . . 22
4.0.4 Interpretation . . . . . 23

**5 Daily Followers and Black Markets** **24**
5.1 Background . . . . . 25
5.2 Features of a Malicious User . . . . . 25
5.3 Vulnerable Features of Twitter . . . . . 28
5.3.1 The Signup Phase . . . . . 28
5.3.2 Experiment Replicating Black Markets . . . . . 30
5.3.3 The Top 21 Accounts . . . . . 31

5.3.4 Experiment and Results . . . . . 33
5.4 FEATURE SET SORTING OF SUSPICIOUS ACCOUNTS . . . . . 33
5.4.1 Analysis and Results . . . . . 33
5.4.2 Nomenclature Analysis . . . . . 33

**6 Relevant Case Study** **36**
6.1 Case Study of 2014 . . . . . 37
6.1.1 August 2014's Political Scope . . . . . 37
6.1.2 Data Gathering . . . . . 37
6.1.3 Experiment on Twitter . . . . . 38

**7 Future Work and Conclusion** **40**
7.1 Third Party Applications and Cyber Crime . . . . . 41
7.2 Solving the issue of Fake Followers . . . . . 41
7.3 Conclusion . . . . . 42

**References** **43**

# Chapter 1

# Introduction

## 1.1 Twitter as an OSN

Twitter is an online social networking service which started in July 2006. Since its inception, Twitter has evolved from a micro-blogging service to a giant pool of interactive dissemination of information. Twitter provides instant access to a plethora of news, opinions, trends and discussions. Twitter's growth has been promising in recent years [23] and it has become a swift carrier of content [3], [2]. As of March 2016, there are over 310 million monthly active users. An estimate volume of Twitter users is around 1.3 billion. 29.2% of social media users in the United States alone are active on Twitter. On average, 500 million tweets are generated every day. Due to its network structure, a user gets visibility over hundreds of other users under the influence of a single tweet. Hence there is a fission effect of information diffusion that envelops a multitude within a short span of time. This social media platform is now being popularly used to interact with other people, exchange information, pool opinions, propagate sentiments, perpetrate narratives, organize networks and bond with people of other geographical territories. Twitter is an active forum of information exchange. Opposed to popular news forums and blogs where people can only passively read or view an news feed but they can not record their remarks about it and share it with other people. Twitter also congregates the most commonly discussed topics in the form of trends. The trends can be classified with respect to geography and location of the user.

There are some key features which make Twitter uique from other major social media platforms. It has an iconic 140 character limit which is short and unfiltered. The character limitation makes it short and simple to use. Unlike Facebook, users do not have to go through a long list of posts and feeds to extract the meaning and context of information. Twitter is also very public in its content provision. Tweets posted by any user is openly available to Twitter users and non users. People who do not use Twitter can still read the timelines of its users. Though they can not participate in the conversation unless they are signed up and logged in. Another reason for Twitter's popularity is the absence of advertisements. There are promoted trends and Tweets but there are no placeholder advertisements that pop up as the user scrolls through. Which is why Twitter is often described as a noiseless social networking site.

Public figures and celebrities mostly prefer Twitter over other social media platforms. Due to its wide reach, extrovert linkage and easy to interact interface, public figures use it for promotion and interaction with fans. In recent presidential elections Donald Trump and Hilary Clinton actively used Twitter to promote their ideas and agendas. President Trump's account became a hot topic for debate on mainstream media and print media. His Tweets stirred hot debates and discussions regarding Mexican wall and remarks regarding immigrants. Twitter also recently added the feature of polling which allows users to get votes of users on a particular issue. This feature became a litmus test for public opinion over various issues. A general trend once Retweeted, reaches out to followers of the Retweeter and then it floats into public domain, engaging users along. Other media icons like movie stars and singers keep their fans updated about their new ventures and also share glimpse of their private life.

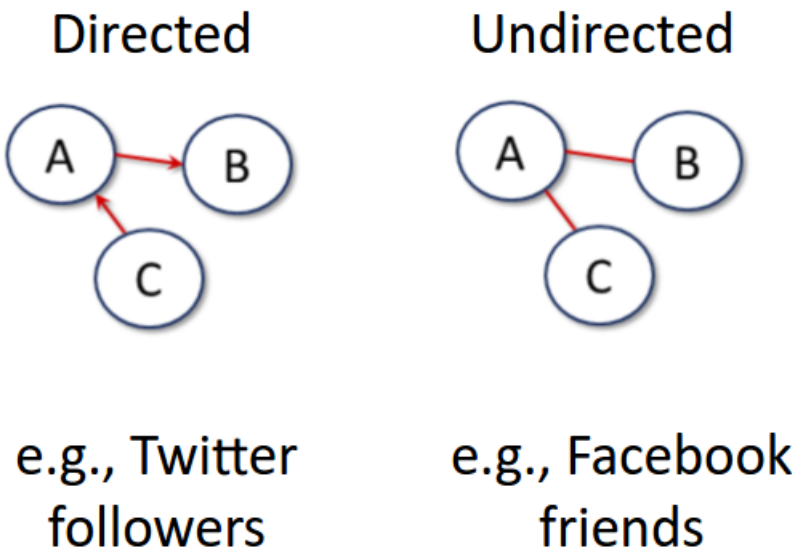



Figure 1.1: Directed link on Twitter

### 1.1 Links on Twitter

Twitter employs directed causal link among its users. A user can follow any other user without consent of the user being followed. Unless the account is made private or the follower is blocked. Figure 1.1 explains the directed link on employed on Twitter. Generally a wide reach can be accomplished based upon massive following. An account with large proportion of followers can expect visibility to a large audience. Since Twitter is a gigantic hub of information, it has a strong propensity of engaging its users and influencing their opinions and sentiments. The directed linkage makes it easier for mesh connectivity of individual users. Its public nature discloses the content exchange to the outside world. Which is why Twitter often gets directly quoted in the form of weblink on other social networking sites.

Figure 1.1 shows directed link on Twitter. This means that C is following A and A is following B. So A can view B's tweets and C can view A's tweets. But on A's timeline, C's tweets will not appear since A is not following C back. Likewise, on B's timeline, A's tweets will not appear since B is not following A back

## 1.2 Engineered Popularity on Twitter

It often comes up that Twitter is sometimes used to promote agendas and directed narratives [13]. We carried out an empirical research to see if there is an organized activity of rhetoric formulation and the domains in which it is frequently applied. It was observed during our experiments that many of the narrative carriers were bots or engineered third party accounts. Some organized groups of humans were also involved in the spread of certain opinions. This gave weight to the belief that there are several collusion networks operate with intentions to spread a directed discourse in public. We used the *CopyCatch* algorithm to detect collusion networks and through these


Figure 1.2: Example of Fake Accounts

colluding groups, we discovered the beneficiary networks which were being supported. Topic modeling of beneficiary networks gave us the insights about popular ideas being ricocheted in the echo chambers of Twitter.

Popularity of a Twitter account is usually measured by the number of followers it has. More followers invariably means a higher liking and fame. Comedian Dan Nainan confessed purchasing fake followers in order to increase fame. After Mitt RomneyâĂŹs account displayed unnatural growth in the number of followers [15], it was discovered that a fake followers market in fact exists and is actually a multi-million dollars industry [1]. Figure 1.2 shows an example of fake accounts on Twitter.

Figure 1.2 shows an example of fake accounts on Twitter. Notice that all the accounts have same name, same biography and location. Since twitter handle has to be unique so all the four accounts have slightly permuted their handles with slight punctuation amendment or an addition of a letter. In coming chapters the actual scheme of creation of such accounts will be elaborated.

## 1.2 Organization

The rest of this thesis is organized as follows. Chapter 2 gives insights towards related work and background research regarding the topic. We also discuss our methodology of acquiring and assembling data set for all the experiments. Chapter 3 will describe the existence and operations of collusion networks. We will present our experiment and results on detecting these networks and monitoring narrative. The second major part of our research comes in Chapter 5 where we outline the problem of "**Daily Followers**" which is an adapted version of **"Fake Bulk Followers"**. We also show our contribution to the old steps of catching fake followers. Chapter 7 gives insights to our future work and conclusion.

# Chapter 2

# Background Research and Dataset

## 2.1 RELATED WORK

Considerable research has been carried out in academia to come up with techniques and classifiers to detect fake followers and malicious activities on online social networks. Twitter in its support article [16] outlines ways to prevent spam, fake following and malicious activities. However Kayode et al. [14] states, âĂIJTwitter lacks an effective spam detection mechanism relying on some rules of thumb to suspend accounts on its network.âĂİ Wang et al. [9] introduced crowd sourcing methods of detecting fake accounts but the analysis does not work well on large data sets. Graph- Based methods and machine learning algorithms have been used by Cao et al. [10], Lee and Kim, 2014 [11], Yang et al. [12] to detect fake accounts based upon feature set aggregate. Yang et al. [12] trained a classifier on features to isolate fake accounts from normal ones. However machine learning techniques are often countered by adaptive attackers who mutate their ways around classifiers.

### 2.1 Classical Approach

The classical approach in detecting collusion networks or systematic group behavior is to use sentimental, temporal and inter-relational features of people in the groups that are collaborating. Ratkiewicz et al. [5] discussed an approach to detect collaborating people by detecting their inter-relations and their posts/tweets. Their work also analyzed text in order to detect sentiments and then using sentiments to catch people with similar views on a topic. Dickerson et al. [8] detected bots on Twitter while using individual sentiment features. The purpose of this work was to detect if bots are more sentimental than real human beings. Ferrara et al. [6] used multiple methods of catching group attacks. Their work also kept into focus the features of individual users in the groups. Our work is directed towards catching accounts as groups of users who are collaborating to support a cluster of other accounts. For this we found *CopyCatch* [7] to be useful. By using *CopyCatch* we were able to spot the organized activity of colluding accounts.

## 2.2 Research on Fake Followers

There are several ways of detecting fake content in the pool of collected data-set. The policy used by Twitter [16] is to keep an eye on skeptical surge in followers count and take measures against it. However the attackers and black markets have diversified their ways to overcome that. Now there is a new feature of **âĂIJDaily FollowersâĂİ** [21] that is in vogue. These markets provide the subscribers with a daily magnitude of followers. In this way a no sudden spike appears in followers count and the account garners followers everyday. This bypasses the thumb rule used by Twitter to detect fake followers. Exploiting the correlation between these **Daily Followers**, we present a graphical and mathematical relation between real and fake followers.

Table 2.1: **Data Acquisition for Collusion Networks**

| Data Characteristics | Magnitude | Description |
|---|---|---|
| Initial Seed | 10 | Profiles selected for analysis |
| Associated Profiles | 50 | All the profiles associated with initial seed |
| Number of Tweets | 160000 | Total number of Tweets obtained from all profil |
| Total number of Retweets | 16 million | Total number of Retweets of all Tweets |
| Cumulative Tweets of all Networks | 1.3 million | All the Tweets of Target Profiles and their collusion networks |
| Cumulative Retweets | 130 million | Aggregate of all the Retweets of every Tweet ob |

Table 2.2: **Data Acquisition for Profile Analysis**

| Data Characteristics | Magnitude | Description |
|---|---|---|
| Initial Seed | 300 | Profiles selected for second part of the study |
| Selected Profiles | 21 | Profiles selected with seemingly unusual growth in followers |
| Total followers of selected profiles | 23 million | Total number of accounts in the pool |
| Associated JSON features obtained | 921 million | Features like status count, creation date, follow |

## 2.3 Acquiring Dataset

Table 1 shows data collection statistics for collusion networks. Table 2 includes data crawled for profile analysis. This data was then for spotting suspicious accounts. Figure 1 displays the data building procedure that we employed. For all our experiments we used the REST APIs and Streaming APIs of Twitter to build dataset. We used eight API keys and Tokens and dedicated a machine to crawl data from August 2016 till December 2016. The machine we used was Intel(R) Core(TM) i5-5200U with 2.20GHz CPU and a memory size of 7897MiB. We used NodeJs to crawl and store data. Once we had the required data set then we converted it into other formats and used it according to the requirements of the experiment.

Here is the list of tools we used to accumulate the data and perform our analysis.

- **NodeJS** for API calls, response parsing and storing results
- Linux command line driven utilities like *gnuplot*
- **AWK** for text processing
- **Python** for visual representation of collusion networks
- **Gephi** to simulate and graphically present network clusters

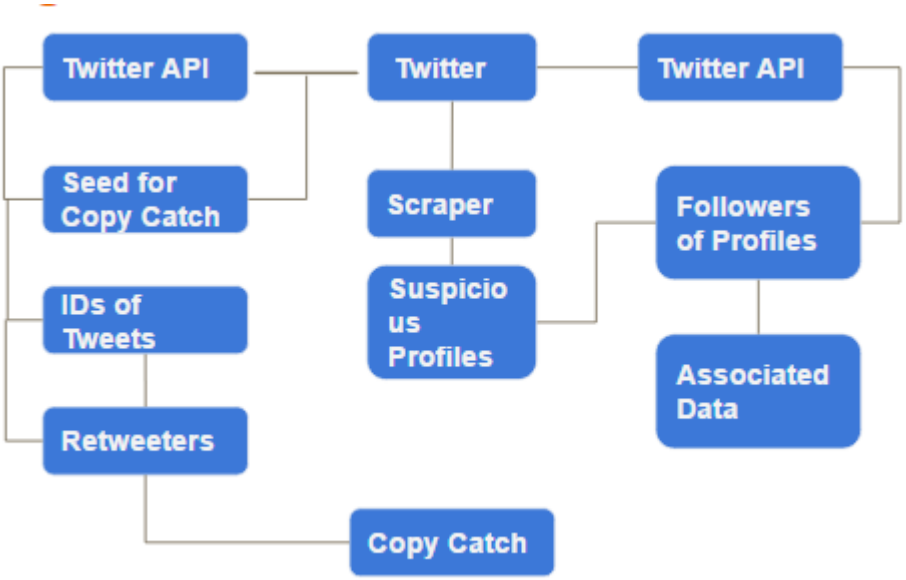



Figure 2.1: Data Assembly

# Chapter 3

# Collusion Networks

## 3.1 COLLUSION NETWORKS on TWITTER

### 3.1 Pretext

The explosion of **Brexit** related posts on Twitter, grabbed attention of many people around the globe. Consequently, we also started analyzing the behavior of users who were *retweeting* posts related to **Brexit**. One of our findings was the behavior of a certain group of users who were *retweeting* both, for and against **Brexit**. We tracked the sources of these tweets to find where they were originated from and found out that the tweets from the same group of users which was *retweeting* for and against **Brexit**, were been originated from a single Twitter app. By exploring this scenario further, we found out that there were more than one group of people whose sources of tweets were same and their online behavior was suspicious.

The decision to use data from Twitter was primarily based on the rate of dissemination of information on Twitter. Where on Facebook, a post is spread very slowly in the social network, due to tighter restrictions on the visibility of the post. Whereas, in case of Twitter, when a tweet is *retweeted*, it is mostly accessible to all the followers of a particular user. Therefore, Twitter provides the most advantageous platform to people with large following to spread information regarding their narrative. Retweets can spread information instantaneously to all the followers of retweeter. So a normal chunk of sentiment can be iteratively made "known" by periodically retweeting it within different clusters of users. This was frequently observed in Presidential Elections of USA in 2016. This information and observation motivated us to analyze ongoing collusion networks on twitter who were engaged iterative Retweeting.

The aim of this section is to analyze and investigate how monetary resources are utilized for promotion in socio-political context. How collusion networks are organized and operated to achieve it. Overall theme is to disclose the phenomenon of planned group activity for a particular agenda. Our first effort was to acquire reasonable data and to process it for further analysis. Furthermore, this work is aimed at categorizing the usage of various techniques used to promote or defame politicians on social media. After categorizing we aimed to identify the people involved in the practice of manipulating social media.

Our work presents various techniques, which help solve this problem:

- **Problem Formulation:** We aim at identifying people using the same set of followers to promote their political agendas.
- **Techniques:** Using various algorithms available today, we identify the people involved in similar activities (promoting or defaming) related to a certain set of profiles.
- **Theoretical Analysis:** Our results show that all the politicians who belong to one political party share the same group of suspicious accounts used to promote all Twitter accounts of this party.

Organized Retweeting is colloquially called Collusion Networks. And we classified the collusion networks in three categories.

1. **Bot Collusion Networks**. Group of in-organic accounts completely run by third party apps. The accounts give complete privilege to the third-party which does Retweeting on their behalf.
2. **Compromised Accounts.** Group of real users who tweet and do normal activity but have given privileges to a third party app. The app on specific actions, performs Retweeting for them. So there is a mixed behavior in their activities
3. **Real Accounts.** Group of real people who Retweet a user or set of users by virtue of their own intent. There is no external third party involved and hence they can not be spotted quickly. However their behavior is planned and systematic.

Besides identifying collusion networks, another aspect of our study includes statistical survey of existence of these collusion networks across Twitter. We find out the segments of social media in which collusion networks are mostly functional. Therefore tracking the key doctrines being intentionally perpetrated. We scrutinized a wide range of profiles and accounts to spot hot agendas being traversed organically. This gave us an intuition about popular consent that is being manufactured across social networks.

### 3.1 Previous Work

The classical approach to detecting this kind of behavior is to use sentimental, temporal and inter-relational features of people in the groups that are collaborating for the deceptive purpose of making politicians appear better or worse.

Ratkiewicz et al. [5] discusses an approach to detect collaborating people by detecting their inter-relations and their posts/tweets. Their work also analyzes text in order to detect sentiments and then uses these sentiments to catch people with similar sentiments on a topic.

Dickerson et al. [8] detects bots on Twitter while using individual sentiment features. The purpose of this work is to detect if bots are more sentimental than real human beings or is it the other way around.

Ferrara et al. [6] uses multiple methods including methods used to catching group attacks. This work does not only focus on catching group attacks but it also keeps into account the features of individual users in the groups, e.g. their political inclinations.

Our work is focused on catching accounts as groups of users acting in favor of a group of accounts; for this we found *CopyCatch* [7] to be useful.

### 3.1 Formulation

In this section, we will start by defining notations used in our methodology. Table 3.1 describes the symbols that we will be using throughout this paper. For further convenience, we will describe the general format of the symbols used. For scalars, we will use italic letters, e.g. $\lambda$ and $Z$. For vectors, we will use bold letters, e.g. $\mathbf{p}$ and

$\mathbf{v}$. For matrices, we will use bold capital letters, e.g. $\mathbf{Q}$ and $\mathbf{U}$. For sets, we will use scripted characters, e.g. $\mathbb{X}$ and $\mathbb{A}$. For indexing of elements in vectors and matrices, we will use subscripts, e.g. $\mathbf{p}_i$ denotes the $i$ th element of $\mathbf{p}$ and $\mathbf{U}_{i,j}$ denotes the element in the $i$ th row and $j$ th column of $\mathbf{U}$. $\mathbf{Q}_{i,*}$ is a vector that contains the elements of the $i$ th row of $\mathbf{Q}$. To use all the elements of a vector in a set, we will use a function $set()$, e.g. $\mathbb{X} = set(\mathbf{v})$. We defined our problem in a way we can approach it practically.

| **Symbols** | **Description** |
|---|---|
| $\mathbf{p}$ | vector of all social media profiles |
| $\mathbb{U}$ | set of all Retweeters common to all profiles in $\mathbf{p}$ |
| $\mathbb{U}_{i,j}$ | set of all Retweeters common in profiles $\mathbf{p}_i$ and $\mathbf{p}_j$ |
| $\mathbb{U}^*$ | set of all $\mathbb{U}_{i,j}$ |
| $\mathbf{O}$ | matrix containing Retweeters with rows corresponding to the profiles in $\mathbf{p}$ |
| $\mathbb{O}_i$ | set of all Retweeters of profile $\mathbf{p}_i$ caught using *CopyCatch* [7]. |
| $set(\mathbf{x})$ | a function that returns a set containing all the elements in an input vector $\mathbf{x}$. |

Table 3.1: **Symbols and their descriptions**

The columns of $\mathbf{O}$ correspond to all the Retweeters that have ever retweeted any one of the profiles in $\mathbf{p}$. This means that if a retweeter has retweeted at least one tweet of a particular $\mathbf{p}_i$ but has never retweeted any tweet of $\mathbf{p}_j$ then the entry in $\mathbf{O}$ corresponding to this retweeter in the $j$ th column will be 'NaN'. Since, $\mathbf{O}_{i,*}$ is a row containing all the Retweeters of $\mathbf{p}_i$. We have to find a set of all the entries in $\mathbf{O}$ such that these entries also exist in $\mathbf{O}_{i,*} \forall i$. We will represent the resultant set by $\mathbb{U}$. Alternatively, we can define $\mathbb{U}$ as a set that contains all the Retweeters that have retweeted all the profiles in $\mathbf{p}$.Defining $\mathbb{U}$, however, does not completely define our problem because the probability of $\mathbb{U}$ being a null set is not guaranteed to be zero in all cases. If $\mathbb{U}$ turns out to be an empty set we will not be able to perform any analysis. Therefore, we find a set of all common frequent Retweeters of $\mathbf{p}_i$ and $\mathbf{p}_j$ and call this set $\mathbb{U}_{i,j}$. A set containing all $\mathbb{U}_{i,j}$ is denoted by $\mathbb{U}^*$. Now, using Figure 3.1, we can narrow down to the definition of our problem. We need to find $\mathbb{U}^*$ so that we can identify those $\mathbb{U}_{i,j}$ that have greater number of elements. Using these elements we can identify those profiles that are connected with a large proportion of elements in $\mathbf{p}$. As a result, we aimed to obtain a graph. This graph contains nodes to represent elements of $\mathbf{p}$ with edges representing the connections among these nodes.

Figure 3.1 gives an example of a Collusion Network and Beneficiary Network. Maroon nodes represent profiles in $\mathbf{p}$. Cyan nodes represent Retweeters in $\mathbb{U}$. Blue nodes represent Retweeters that are unique to each profile. The blue nodes are collusion networks associated with each maroon node. Cyan nodes are also the part of collusion network with blue nodes but they support multiple nodes at the same time. In other words they are part of more than one collusion networks. All profiles in $\mathbf{p}$. So from

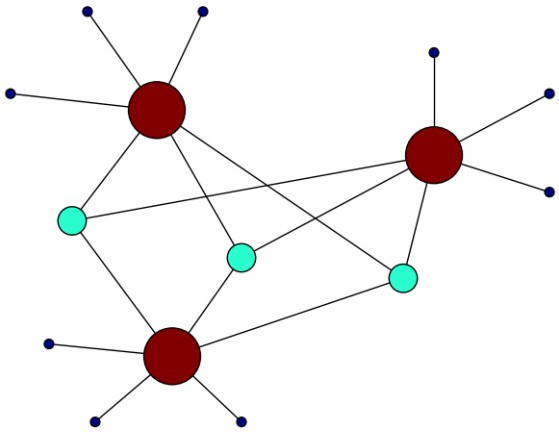


Figure 3.1: Collusion Network Example

the perspective of cyan nodes $\mathbb{U}$ the maroon nodes $\mathbf{p}$ form a beneficiary network. Tweets of $\mathbf{p}$ reflect the common narrative which is being propagated by $\mathbb{U}$.

## 3.2 EXPERIMENT on COLLUSION NETWORKS and RESULTS

The following steps give an overview of our methodology of tracing collusion networks and beneficiary networks.

1. Select initial seed of profiles as mentioned in Figure 2.1
2. Isolate the bots and compromised accounts by using source of their retweets. If it is a third party app, then these accounts are to be isolated and put in the category of automated collusion networks
3. For the organic users selected, use *CopyCatch* [7] to obtain a set of users that tweet atleast $n$ tweets along with $m-1$ other users within a time window of $\Delta t$. This set of users form a set $\mathbb{M}$. When the algorithm converges, we have a collusion network that operates in a lockstep behavior.
4. To discover the beneficiary network, obtain a set of $k$ profiles that are most commonly retweeted by all the users in $\mathbb{M}$. The elements of this set, when put in a vector will form $\mathbf{p}$.
5. For each profile in $\mathbf{p}$, find a set of $r$ top Retweeters. We represent each of these sets as $\mathbb{P}_i^*$, where $i$ corresponds to the $i$ profile in $\mathbf{p}$.
6. For all $\mathbb{P}_i^*$ obtained in the previous step, find profiles that these Retweeters tweet more frequently. We can call these profiles $\mathbb{L}_i$, where $i$ corresponds to the $i$ th profile in $\mathbf{p}$.
7. Start plotting a graph, where all elements of $\mathbf{p}$ are the initial nodes in the graph. Then for each $\mathbf{p}_i$ make edges from its node to all nodes in $\mathbb{L}_i$, keeping

under consideration that some or most of these $\mathbb{L}_i$ might be the same nodes that represent elements in $\mathbf{p}$. The size of the nodes is a function of the number of edges connecting this node to other nodes.

### 3.2 Use of *CopyCatch*

To start our analysis we needed a list of Twitter users with anomalous behavior related to $q$. We chose *CopyCatch* for obtaining this list of users because it detects *lockstep* behavior while keeping *temporal* features into account.

*CopyCatch* needs the following parameters as inputs. There are other parameters that need to be set; however, for details of those parameters, please refer to the original publication [7].

- $n$ the minimum number of retweets that anomalous users have to do as a group, in order to be considered anomalous.
- $m$ the minimum number of anomalous users in a group, in order to consider that group of users anomalous.
- $\Delta t$ the time window within which atleast $m$ users from the anomalous group have to retweet atleast $n$ tweets.

The problem that *CopyCatch* solves is formulated as finding sub-bipartite graphs in a bipartite graph of higher order. It fitted our purpose of finding an initial list of retweeters of $q$ because our problem of finding initial set of retweeters could be mapped exactly as finding sub-bipartite graphs. The literature on finding sub-bipartite graphs focuses on co-clustering; *CopyCatch*, however, is an alternative to co-clustering that achieves the same goal.

*CopyCatch* operates on two matrices $\mathbf{L}$ and $\mathbf{I}$. $\mathbf{L}$ has rows corresponding to all the retweeters that ever retweeted $q$ and the columns corresponds to all the tweets that $q$ ever tweeted. The elements in $\mathbf{L}$ represent the time at which retweeters retweeted a certain tweet. The rows and columns of $\mathbf{I}$, also correspond to retweeters and tweets respectively, however, the binary elements of $\mathbf{I}$ represent whether a particular retweeter ever retweeted the corresponding tweet or not.

The solution is approached by defining this as an optimization problem. If a user retweets a certain number of tweets within a specific time window then the tweets that this user has retweeted form a subspace, $\mathbb{X}$, in which the retweeter is defined as a point, $\mathbf{c}_k$. The average $\mathbf{c}_k$ of all the users that have retweeted this particular tweet within a specific time window is defined as $\mathbf{c}$. The optimization focuses on updating $\mathbf{c}$ and $\mathbb{X}$. Proof of convergence can be found in the original work [7].

Finally, we can output a set of all the suspicious retweeters. As defined earlier we call this set $\mathbb{M}$.

### 3.2 Detecting Beneficiary Network

$\mathbf{p}$ is a vector that contains profiles related to a common theme or agenda. We can obtain this by manipulating $\mathbb{M}$ that was obtained as a result in the last subsection.

Maintain a vector of profiles, $\mathbf{p}$, that have been retweeted by every retweeter in $\mathbb{M}$. Iterating through each element of $\mathbb{M}$ we can increase the count corresponding to the profiles in $\mathbf{p}$. Finally we will sort $\mathbf{p}$ according to the retweet count obtained while iterating through each element of $\mathbb{M}$. The elements of $\mathbf{p}$ are the key focus of our investigation, since $\mathbf{p}$ is a vector of all the profiles that have a common agenda. For each profile in $\mathbf{p}$, we obtain a set of top Retweeters, i.e. a set $\mathbb{P}_i^*$ corresponding to $\mathbf{p}_i$. All $\mathbb{P}_i^*$ are sorted by the number of retweets that the respective retweeter has done. The set of all $\mathbb{P}_i^*$ gives us all the Retweeters that are retweeting the profiles in $\mathbf{p}$. The activities of these $\mathbb{P}_i^*$ are critical to the investigation, since these are the Retweeters that are spreading the narrative to a large number of audience. The impact that the Retweeters have is because of their retweets reaching out to large number of people, either directly or through retweets by their followers. Therefore, after acquiring all $\mathbb{P}_i^*$ in the previous paragraph, we find out who they are retweeting the most. To do this, we acquire retweets from the profiles of each of these Retweeters. Once we have these retweets we find the profile which originally tweeted the content of the retweet. All the profiles acquired in such manner are put into a set $\mathbb{L}_i$ where $i$ is used for representing correspondence to the profiles in $\mathbf{p}$.In short, we can define $\mathbb{L}_i$ as the list of the profiles that are retweeted most frequently by the users that also retweet $\mathbf{p}_i$ frequently

### 3.2 Tools

#### Acquisition and processing of data

We are using **NodeJS** to acquire data from Twitter. The data acquired by Twitter API is in JSON by default. Since **NodeJS** is built on Google's open source high-performance JavaScript engine, V8, it has a global object JSON which can be used to parse and stringify JSON strings. Even though the handling of JSON is synchronous, the overall asynchronous nature of **NodeJS** helps us acquire data through API calls much more quickly as compared to synchronous libraries available on other platforms.

#### *CopyCatch* and Data Representation

**Python** provides an excellent platform that has native support for set operations and provides abundant resources for matrix operations. We are using ***numpy*** because of its built-in support for numerous matrix operations and a large community that is involved in development of applications that utilize ***numpy***. ***NetworkX*** is built around visualizing network data and perform network analysis; due to its design and ease of use we selected this library for representation of network of politicians promoting a common agenda. Whereas ***Plotly*** helps us to plot statistics related to our results.

### 3.2 Interpretation

*NOTE: Identity of Political figures has been masked and user discretion is maintained. However for the account under scrutiny here, the name has been replaced by Alice.*

#### Selecting initial profile

We selected the account of Alice due to a recent surge in its followers and Retweets count. Alice's party narrative on social media was very shabby over years. However, very recently the cyber environment suddenly changed in Alice's favor. Public is taking it with a pinch of salt while we decided to permeate into the technical aspect of it. We observed that multiple Twitter profiles gained a huge number of followers during a short period of time. However, in particular the number of followers of Alice's Twitter profile increased from $\sim 15,000$ to $\sim 1.5 million$ within a period of 3 months. This motivated us to select the Twitter profile of Alice as the starting point of our investigation.

#### Use of *CopyCatch*

We analyzed the retweets on Alice's Twitter profile using *CopyCatch* and found a group of people that retweeted most of the tweets within a short time window. These were the set of people that whose behavior does not follow standard behavior of a Twitter user. These people only retweeted Alice's and other profiles related to Alice's party.

#### Detecting key profiles

Using the set of profiles we obtained using *CopyCatch*, we obtained a set of profiles that were being retweeted the most. The following profiles were present in this set.

- Alice
  (https://Twitter.com/MaryamNSharif)
- Zeshan Malick
  (https://Twitter.com/ZeshanMalick)
- Govt Of The Punjab
  (https://Twitter.com/GovtOfPunjab)
- M. Shams-Uz-Zaman
  (https://Twitter.com/MShamsZ)
- (https://Twitter.com/AadiiRoy)
- Shehbaz Sharif
  (https://Twitter.com/CMShehbaz)
- PML(N)
  (https://Twitter.com/pmln_org)
- Geo News Urdu
  (https://Twitter.com/geonews_urdu)

The *common* anomalous retweeters relate these profiles. In this way we obtain a set of profiles are using a common pool of followers to promote their tweets.

### Top retweeters of key profiles

To strengthen the connections between these key profiles, we made a list their followers and sorted them in order of the number of retweets that each follower does for the corresponding key profile. These top retweeters are the profiles that can help us detect the links between the top level key profiles.

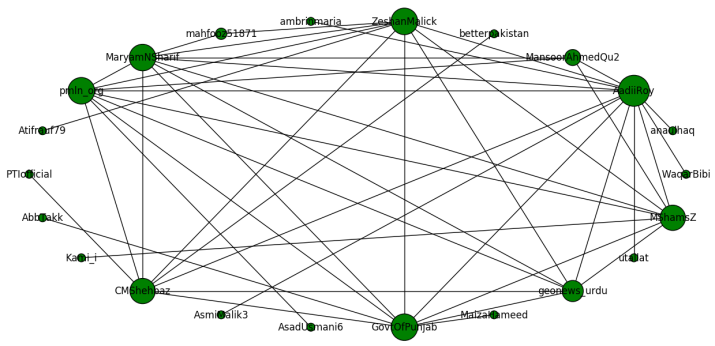
### Retweets of top retweeters

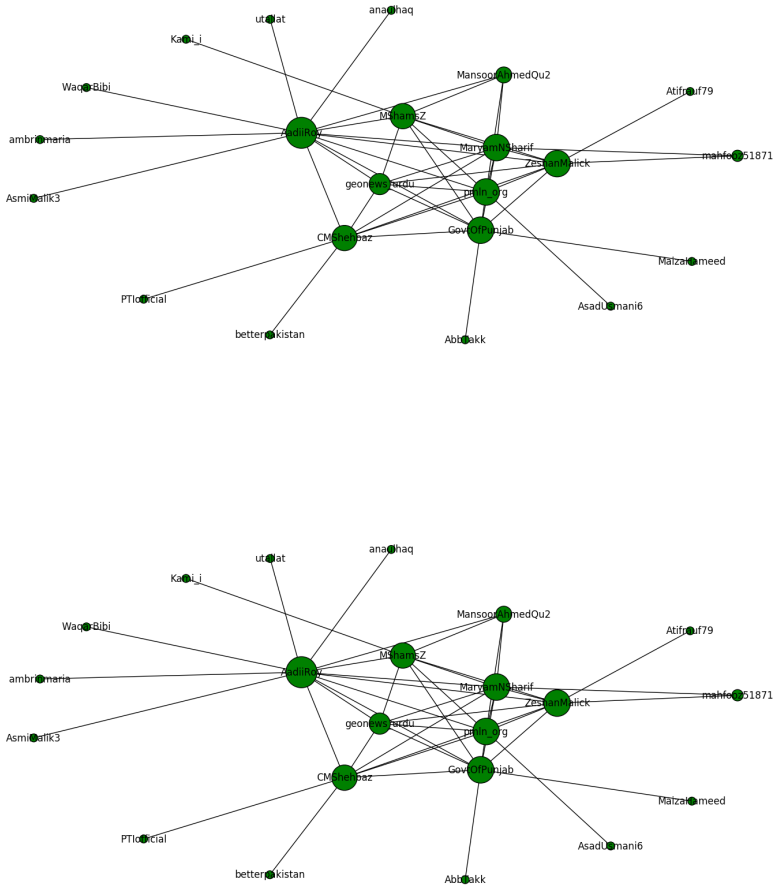
Having obtained a list for each key profile we begin making a graph that links each key profile if the top retweeter of one profile also frequently retweets another key profile. The resultant graph shows that the key profiles that we detected have the majority of their tweets retweeted by a common pool of users, regardless of the content of the tweets.

This is an indication that these key profiles are involved in obtaining *anomalous* retweet counts.

### Visualizing the results

The following graphs are represent different configurations of the same network of key profiles.









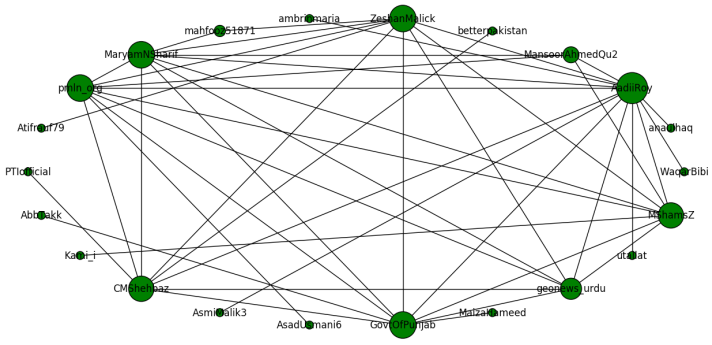



Figure 3.2: Final Results

The bigger nodes are the key profiles and smaller nodes represent other profiles that get retweets by the top retweeters of these key profiles.

# Chapter 4

# Monitoring Narrative

Table 4.1: **Results of Collusion Network**

| Accounts | Bots | Key Third Party Apps | Organic Networks | Results of CopyCatch |
|---|---|---|---|---|
| ICC | ✓ | 20cmJothipala Bot, t2r app 3, TagtheBirdbot, Competitions449 and 12 more | X | 0 suspicious user found |
| CNN | ✓ | 40 cm FishesPetMag, People To Retweet, nodeAppl12, Sapien_Bot | | |
| Dabr, news_genertor, lorelei2016, Twicca, amtrendbot, blog1b1, retweet agent And 179 more | ✓ | 184 Suspicious users found | | |
| FOX | ✓ | 40 cm Metaltrack RT, Tabtter Free, YoruFukurou, QxNews-python, Bird Nest | | |
| RoundTeam, Eustace and 84 more | ✓ | 229 suspicious users found | | |
| FIFA | ✓ | Indiansuperbot, Senegal bot, LTP14, queuestreamer7 and 24 more | X | 0 Suspicious users found |
| Democrats | ✓ | 40 cm TruffalaSeedBot Memory-Bot EhWiederBeste second test 420 | | |
| cyborgthemanrobot Labour North Robin's NewsWire Retweets and 254 more | ✓ | 229 suspicious users found | | |
| Republicans | ✓ | Cloudhopper, 360Vid Bot, Perioscope, RoundTeam and 12 more | ✓ | 113 users found |

## 4.0 Monitoring Narrative

To monitor the spread of discourse through collusion networks, we selected profiles from diverse domains. Some notable ones were **ICC**, **FIFA**, **The Democrats**, **House Republicans**, **Fox News**, **CNN**. We picked all the associated verified accounts relevant to these key accounts and also selected personal profiles of politicians, celebrities, singers, social activists, sportsmen and journalists. Based upon scale and size of collusion or beneficiary networks, we made an estimate of the popular narratives being promoted artificially on social media. Our results show that in the field of sports there were no organic collusion networks. There were some third party apps but their contribution was insignificant. This can be verified by simple eyeballing of their accounts. For example **ICC**'s official account has more than 4.81 million followers but its Tweets are hardly ever retweeted. So once CopyCatch was run over these followers there were no suspicious accounts.

In contrast to that every political wing and news forum had employed collusion networks for promotion of their tweets There were more than 3000 unique third party apps active in Retweeting these accounts. Most commonly used third party apps for scheduled Retweeting are *Metaltrack RT, Tweetbot for iÎ§S, Romaxbot, Hootsuite, Tweetbot for Mac, Tweetlogix, PiedCro Bot2, YoruFukurou, Talon, Buffer, SillyConBot, Falcon Pro 2015, bostonlocalnews00, syriabo, amtrendbot, Visual Decolonization Tweet Bot, LockHimUP21, lorelei2016, TwitterKingAPI, My awesome tweeting bot, SapienBot, news_generator, Deputy Bot, amtrendbot, aut0retweet, wwics app, Sprout Social, Sad Trump Club, blog1b1, retweet agent, kyply_bot*

Table IV shows that collusion networks of real people were most commonly found in political and media camps. Every political sphere had organized people who were disseminating the centrifuge of their ideology. They were also generating trends in an organized pattern and Retweeting trends of their collusion networks. The Democrat wing is perhaps more rigorous in their pursuits since 229 suspicious accounts were caught by CopyCatch. The Republican faction however is more dependent on Bot collusion network rather than organic users but their bot networks and organic networks are much smaller compared to Democrats.

Next to political section, the news cells have very organized networks of bots and real people who Retweet them frequently. Every news cell has acquired this technique and they prefer organic collusion networks rather than bots Table IV. the sports division is oblivious to any sort of colluding network. It is to be noted that organic networks are much more useful compared to bots because suspicious users caught by CopyCatch have more following than compromised accounts or bot accounts. Since they have more following, so they have a wider reach and unlike bots, they are able to engage

other users with replies and Direct Messages.

The Table IV shows that in Politics, Democrats are more enthusiastic in reaching out to people and spread their agendas. To uncover their major themes, we did **Latent Dirichlet Allocation** [20] topic modelling on the Tweet texts. The tweets most frequently retweeted by collusion networks were pooled in one array. LDA is a machine learning algorithm that extracts topics and their related keywords from a set of sentences.The algorithm employs probabilistic model for detection of specified topics. Figure 4.1 shows sample output of the LDA feature set analysis done on Democrats. It is an array of Javascript objects with keys being the topics and value being the corresponding probability. For more details pertaining to LDA, consult [20].

### 4.0 LDA Example

Assume that we have a set of following words.

- President **Trump** asks for ban on **immigrants**
- *Obamacare* and *health* issues will be addressed by the democrats
- President **Trump** wants good relations with *Russia*

Latent Dirichlet Allocation (LDA) is an algorithm that intelligently discovers the existence of certain topics in a document. Provided the sentences above, LDA might classify the bold words under the Topic F, which we might label as âĂIJtrumpâĂIJ. Likewise the italic words will be classified under a different Topic P, which we might label as âĂIJpresidentâĂIJ. LDA defines each topic as a bag of words and assigns them weight through an algorithmic iteration.

There are two benefits of LDA defining topics on word level

1. The content spread of each word can be found over sentences.
    - Sentence 1: 100% Topic F
    - Sentence 2: 100% Topic P
    - Sentence 3: 50% Topic P and 50% Topic F
2. We can derive the proportions that each word constitutes in given topics.

LDA achieves topic modeling mentioned above in three major steps.

- **Step 1:**

  The user tells the algorithm how many topics to assign to the given document. This can be done by regression and trial or if there is a prior knowledge regrading the information at hand. Often previous work regarding the topic gives certain insights on the number of topics. But if they are not known, then trial and error can be performed. In trial and error method a limit can be set on to input parameters which produce the maximum statistical output. Or in other words the number of topics which yield the maximum probability. Sometimes topics

can just be inferred from plain study of the documents. In the example shown above the topics can easily be estimated.

- **Step 2:**

  The algorithm will assign every word to a temporary topic. Topic assignments are temporary as they will be updated in Step 3. Temporary topics are assigned to each word in a semi-random manner (according to a Dirichlet distribution, to be exact). This also means that if a word appears twice, each word may be assigned to different topics. Initially there is nothing to worry about this issue. Eventually when the algorithm converges, this problem is automatically settled. The algorithm iteratively updates these topics, computes probabilities and assigns the most probabilistic topic to every word. Note that in analyzing actual documents, function words and stop words (e.g. "a", "an", âĂIJtheâĂİ, âĂIJandâĂİ, âĂIJmyâĂİ) are removed and not assigned to any topics. Stop words do not contribute to the overall theme of any topic. Mostly all the topics assigned to a list of words or documents are nouns.

- **Step 3:**

  The algorithm will check and update topic assignments, looping through each word in every document. For each word, its topic assignment is updated based on two criteria:

  1. How prevalent is that word across topics?

  2. How prevalent are topics in the document?

  Weighing conclusions from the two criteria, we would assign particular topic to a word. Using the estimates a probabily of occurrence will be assigned to a word. This would display the certainity with which LDA modal assumes that the word belongs to that topic.The process of checking topic assignment is repeated for each word in every document, cycling through the entire collection of documents multiple times. This iterative updating is the key feature of LDA that generates a final solution with coherent topics

### 4.0 Sample LDA Result

Figure 4.1 shows a sample output for **Latent Dirichlet Allocation** processed Tweets of our target accounts.

```
[ { term: 'thedemocrats', probability: 0.095 },
  { term: 'trump', probability: 0.033 } ],

[ { term: 'obamacare', probability: 0.091 },
  { term: 'trump', probability: 0.07 },
  { term: 'president', probability: 0.055 },
  { term: 'health', probability: 0.023 } ],

[ { term: 'russia', probability: 0.09 },
  { term: 'muslim', probability: 0.032 },
  { term: 'ban', probability: 0.032 },
  { term: 'gop', probability: 0.031 },
  { term: 'immigrants', probability: 0.029 } ]
```

Figure 4.1: Sample LDA Result

### 4.0 Interpretation

The result of running LDA analysis on the entire text corpus of Tweets and replies shows the engineered information being populated among normal users. The sample result shows that the collusion networks of Democrats were aggressively Retweeting the Tweets involving terms like **Ban**, **Immigrants**, **Health** etc. In a broader view, a group of accounts were dedicated to spread the given agendas and create the effect of an echo chamber for the users who follow them.

# Chapter 5

# Daily Followers and Black Markets

## 5.1 Background

The famous app âĂIJStatusPeopleâĂİ makes an estimate of fake followers on Twitter. According to the app, 47% of **Barack Obama**'s (85.4 million) followers are fake. The same application gives 24% estimated fake following of **John McCain**. When the news of Fake Followers made its way to the mainstream media, Twitter took some measures to apprehend malicious accounts. If a certain account suddenly showed humongous growth in followers count, Twitter removed those followers. A lot of research is done in this field which is mentioned in Section 5.3. To circumvent detection, the Black Markets adapted to new ways of providing Fake Followers. Instead of giving bulk of followers at once, they started giving chunks of *Daily Followers* [21]. To study this phenomenon, we carried out a series of experiments. Our work builds upon the previous work done in the field of spotting fake followers 5.3 but it addresses a unique problem of adapted strategy of Black Markets. Due to features in the sign-up phase of Twitter, a number of accounts can be made using an automated script These accounts can be operated without verification of email and black markets exploit these features.

## 5.2 Features of a Malicious User

For a normal account the rise in the number of followers has a non-deterministic behavior. As the sample space of days increases, the deterministic probability of uniformity decreases. The followers count depends on a number of factors which include popularity, activity, usage, accessibility and intractability. The propensity of growth in your followers is directly related to the number of people who know you. But irrespective of whether you are extremely popular or not, the growth in your followers is never precisely deterministic. For example if a particular account is famous and on a certain day it gets 4000 followers, then the number of followers should vary in the days to follow to reflect an organic behavior. If the gain in followers remains around 4000 for next week or month then the growth is unnatural.

To reflect this mathematically, correlation functions can be used to determine a bot process or a human process in a collective activity. In graphical models, correlation coefficients determine the behavior of a particular process indexed with time, with respect to a similar process in the same time period. Some common correlation models used are Pearson Correlation and Spearman's rank. For simplicity's sake, simple correlation coefficient can also be used to obtain desired results.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Where $n$ is the sample size, and $x$ & $y$ are sample values from two disjoint populations.

Our intuition was that there are certain accounts that subscribe to underground markets and gain followers. Various websites are available on the web which offer numerous plans for getting daily or bulk followers. So for our first experiment we looked at a pool of accounts and narrowed the ones whose followers' growth was

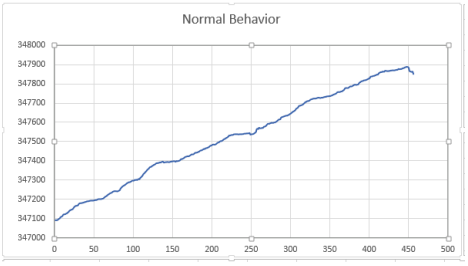
unusually high. We then monitored their followers count using Twitter's REST API. After every 15 minutes, the accounts returned their followers count and we saved that count.

So in a day we got 96 discreet values of the followers count. We did this for over a week and isolated the accounts which represented normal pattern of growth and the ones which displayed aberrant behavior. For the test experiment we looked into the top accounts of the country. Not surprisingly many of them demonstrated unusual activity. Figure 5.1a shows the activity of a normal account with a constant growth in followers. Figure 5.1b shows the activity of one of the malicious accounts. For over 500 discreet values both the graphs are showing an increasing trend. There is a significant overall variation in the normal account and the behaviour is not entirely deterministic. This reflects normal behavior. In contrast, the malicious account showed a deterministic constant growth in the followers count. There were smooth transitions, predictive growth and no variations. This reflects bot behavior. We tested and plotted one normal account behavior against three malicious account behaviors. Figure 5.1c and Figure 5.1d show the contrast between the normal account and two malicious accounts. Figure 5 and Figure 6 show the relationship in the behavior of malicious accounts. It is apparent from the figures that the users seem highly related. However, to display the similarity in a more figurative way, we took the maximum value of each figure's data points and normalized all the data points by that maximum value such that all of them almost converged to 1. Lastly, we plotted them in 5.1h. The three lower curves belong to suspicious accounts and can be seen as almost wrapped together while the upper line represents the normal account.

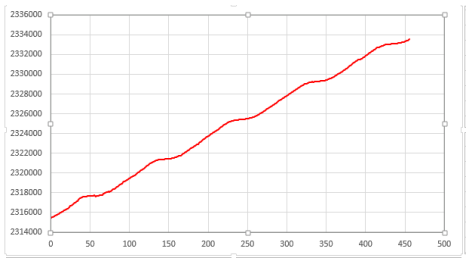
The mathematical way to differentiate between a human behavior and a bot behavior is by finding the cross correlation. The cross correlation coefficient tells the degree of relationship between two curves. If two individuals have subscribed to the same underground market and are gaining followers, then the growth patterns of their followers count will have significant similarity and their cross correlation factor will be high. It is also expected by a bot behavior to repeat itself continuously. However, a normal account's relationship with its behavior in the past or future is independent and less likely to be related. Our first experiment differentiates the normal behavior accounts from the deviant ones. Later in the paper we build the case around the widespread impacts of these accounts' unusual activities.

For cross correlation we picked up a normal account as the test and three malicious accounts as the comparison tool. The idea is to show that there is a strong relationship in the behavior of malicious accounts and disjoint behavior with that of a normal account. The accounts were checked to collect 500 discreet followers count every 15 minutes. Since all the graphs show an increasing trend, the correlation coefficient was overall high. It was specifically higher within malicious accounts. So to highlight the difference in a visible manner, we devised a formula and built a matrix derived from it. Figure 2 shows this matrix.

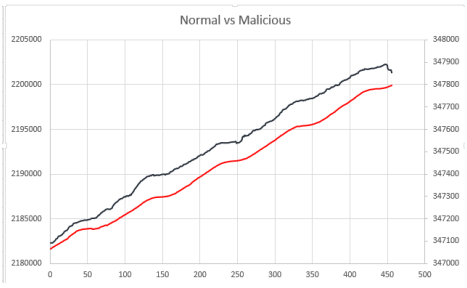
The formula extends the correlation factor in a way that a binary assignment of highly correlated and less correlated datasets becomes possible. Accounts with a strong correlation factor get an assignment of 1 and less correlated data sets are assigned 0. This is done just to make the visualization and classification simpler. Figure 2 shows that normal account has factor 1 correlation with itself and 0 with malicious accounts.



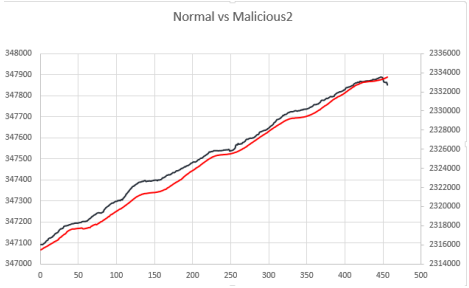

(a) Normal Account



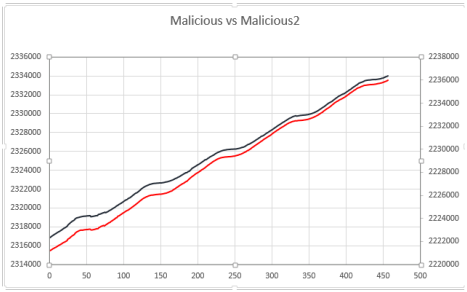
(b) Malicious



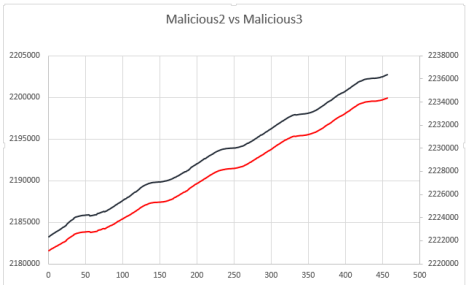

(c) Normal vs Malicious1



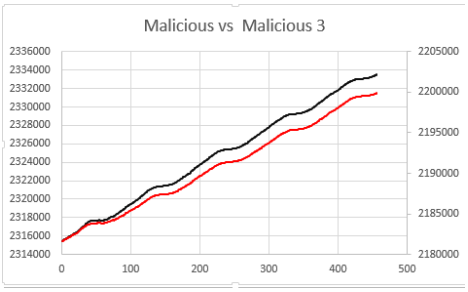

(d) Normal vs Malicious 2



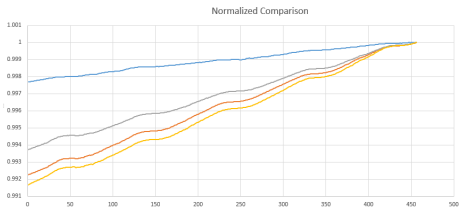

(e) Malicious vs Malicious 2




(f) Malicious 2 vs Malicious 3




(g) Malicious vs Malicious3




(h) Normalized Pattern

Figure 5.1: Normal vs Malicious Accounts

On the other hand, malicious accounts have factor 1 correlation with themselves and other malicious accounts and factor 0 correlation with normal account.

$$\mathbf{L_{i,j}} = correl(\mathbf{X_i}, \mathbf{X_j})$$

$$\mathbf{D_{i,j}} = \frac{sign(1 - round(1000 \times (1 - L_{i,j}))) + 1}{2}$$

In this case, $D_{i,j}$ acts as a step-function $0 \longleftrightarrow 1$ with value 1 when $X_i$ and $X_j$ show correlation $\geq 0.996$ and 0 in all other cases.

To determine the repetitive pattern of these malicious accounts, we took the same

| | Normal Accoun | Malicious 1 | Malicious 2 | Malicious 3 |
|---|---|---|---|---|
| Normal Accoun | 1 | 0 | 0 | 0 |
| Malicious 1 | 0 | 1 | 1 | 1 |
| Malicious 2 | 0 | 1 | 1 | 1 |
| Malicious 3 | 0 | 1 | 1 | 1 |

Figure 5.2: Correlation Matrix

data set with 500 points and split it. We plotted curves with 96 data points for the first set and then repeated it for the remaining points. Hence we got curves showing an accountâĂŹs behavior in terms of sequence of days. First 96 points corresponding to day 1, the next 96 points corresponding to day 2 and so on. The correlation between day 1 âĂę day N showed the overall correlation in the behavior of followersâĂŹ growth. Our experiment showed that bot behavior generally tends to have high correlation factor in the sequence of days due to deterministic machine pattern, while the normal human behavior tends to vary from one day to the next. .

## 5.3 Vulnerable Features of Twitter

Knowing that there are black-hat markets that provide fake followers to their subscribers, there was one more peculiar thing about the whole idea. Initially, it was widely assumed that these underground markets have a large pool of organic accounts and upon requests, these accounts may be used to follow the target account. However, this would have resulted in a sharp rise in the number of followers of a target account, instead of a regular pattern. An immediate growth in followers can easily be spotted by Twitter and fake followers were often removed due to an unusual spike. To bypass this detection, underground markets use an alternate method where theyprovide daily follo,wers in a certain proportion. As a result, the growth is not sudden and the suspicion regarding fake followers is often reduced. This circumvention technique involves the creation of daily accounts by exploiting some structural flaws in Twitter's framework.

On manually inspecting the fake accounts created daily, we observed that these accounts are not organic in nature and are being created everyday in thousands. To probe further, we looked into some top websites that provide these services and we noticed the scheme of daily followers in the list. It was slightly expensive than organic bulk followers but was readily available on payment. This led us to believe that several accounts are created daily to follow a set of targets. While attempting to create similar accounts manually, we came across some security flaws in the sign-up procedure for Twitter.

### 5.3 The Signup Phase

If a legitimate user has signed up for Twitter using âĂIJtest@gmail.comâĂİ as their email address, Twitter will not allow another account to be set up over the same email ID. However, this feature alone is not as secure as it may seem. One possible

circumvention method is This process can easily be circumvented by adding a full-stop or a plus symbol anywhere before the '@' character. So, if âĂIJtest@gmail.comâĂİ is changed to âĂIJt.est@gmail.comâĂİ or âĂIJtest+@gmail.comâĂİ, Twitter will allow the creation of an account against the tampered email ID. However, the verification link will be sent to the original email address. In this way, a number of accounts can be created over the same email ID.

Another way of signing up is by using a ghost email ID, one that does not exist e.g âĂIJthisemailcannotpossibilyexist11987462728626@gmail.comâĂİ. If email the account does not exist, Twitter will inform the user that a verification email has been sent to the email ID regardless. This appears to be a rather easy way of creating accounts over nonexistent email IDs and permuting them with special characters mentioned previously, to create as many accounts as required. This also raises questions about how many accounts on Twitter actually belong to the people who own a particular email ID.5.3a shows that when a malicious user signs with test@gmail.com", they are not allowed to proceed further. 5.3b indicated that when the user punctuates the same email string with a dot t.est@gmail.com" they are granted access, indicated by a blue tick on the right corner. Once the first phase is passed, the user ends up on a page where they are asked to verify their phone number. Unlike other social media forums, several accounts on Twitter can be registered over the same phone number. Even more alarming is the fact that one can easily bypass this phase by clicking the skip button. 5.5a illustrates this.

By clicking the skip button the malicious user bypasses the confirmation phase and goes onto the page of selection of username. Even that can be skipped by clicking skip button 5.5l. By this time an account is created and a user can skip the next phases of signing up and can move onto homepage. However to a âĂIJDaily FollowersâĂİ service provider, the actual activity begins here. The final screen 5.5c shows 21 most popular accounts based upon your location. On that screen user can also add other target accounts that it intends to follow. If the user is not interested in following some of the 21 accounts recommended by Twitter, then he can uncheck them and continue.

We call this âĂIJTop 21âĂİ and we will later show our analysis of these top 21 accounts. The identities of accounts are not shown for privacy concerns. Now here are some possibilities associated with this peculiar behavior.

- If a user follows those 21 accounts then he becomes a common follower to all
- If any of the 21 account subscribes to the service of daily followers, he invariably provides followers to the other 20 accounts.
- If a normal user subscribes to such a service, then based upon the location of their service provider, they become a beneficiary to 21 recommended accounts of that location.
- âĂIJDaily FollowersâĂİ providers can easily automate these steps and generate accounts. Since we deciphered this theory, we also replicated this upon a test account.

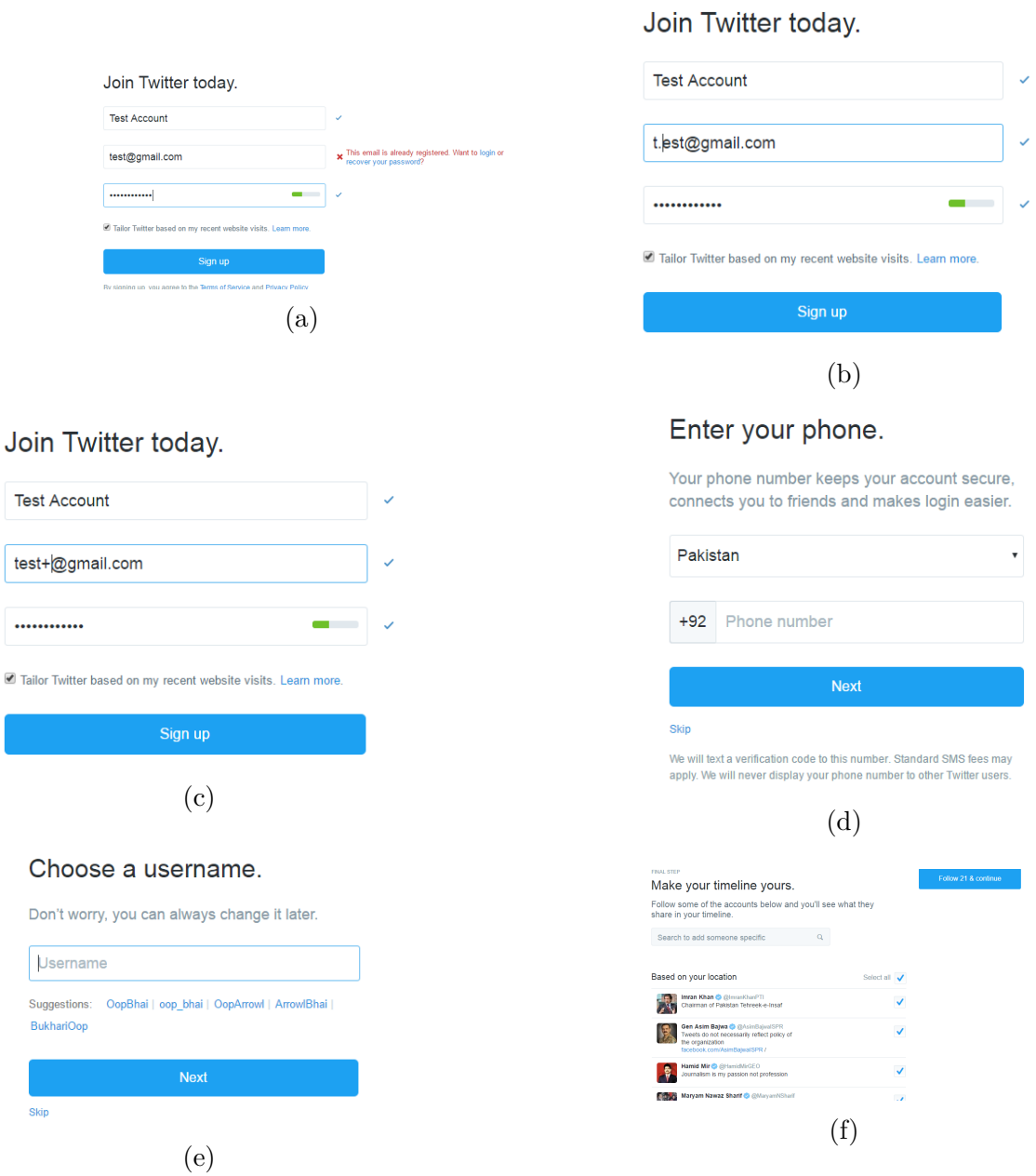



Figure 5.3: Twitter Signup

## 5.3 Experiment Replicating Black Markets

We created a web automation script in a headless browser and by using a pool of random names and ghost email IDs, we created about 400 accounts and followed all the above mentioned steps. We unchecked all the accounts in the top 21 and followed the target accounts. Upon the completion of the experiment, the accounts were subsequently deactivated. Since the 400 accounts were not following any other account, so the activity did not spread to other accounts and was instantly terminated. However this gave us the key insights about how such providers of followers work. We believe that if an account is created upon a permuted email of a valid account and some serious malicious activity is performed (cyber crime, harassment and violent threats, for example), then the valid users may ultimately suffer. Assuming the credibility and

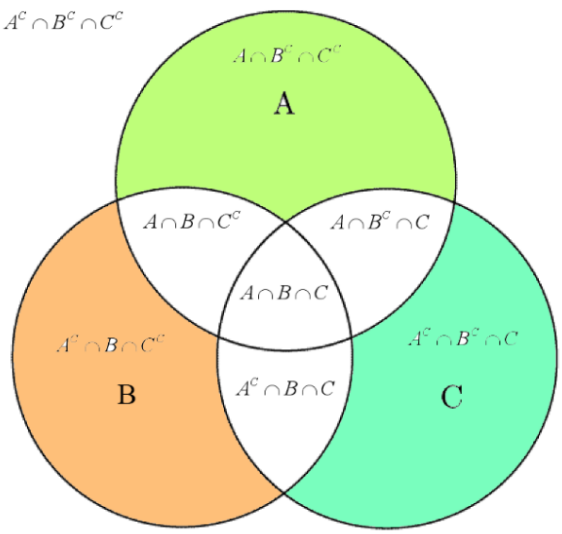



Figure 5.4: Sum of mutually exclusive followers

impact of Twitter, this situation must be taken very seriously and dealt with using appropriate measures.

### 5.3 The Top 21 Accounts

As mentioned earlier, the top 21 includes 21 accounts based upon geographical location and popularity measure attached to them. They can also be modeled by mentioning interests on a screen before the final screen. Provided that in any given club of 21 if a user has subscribed to a service of followers, other twenty would be the indispensable recipients of those followers. Or in another scenario, if some of those accounts have subscribed to the same service then their followers would be following the same accounts. If the provision of followers is not selective, then mutual followers of all 21 will be reasonably high. If the provision was selective, then percentage of of mutual followers would show spikes in mutuality.

We devised an algorithm that converges the following count of the top 21 accounts in a way to show their mutual followers. One way was to crawl a set of followers of all 21 accounts and find out the intersection of each account with every other account. This turned out to be a laborious and tiresome task, since all possible intersections of 21 accounts would have been many. We propose the Followers Distribution Algorithm that solves the overall issue of finding distributed mutuality of accounts. Our algorithm shows the spread of bulk followers in these 21 accounts. We make use of iterative object key value pairs to calculate the number of followers, following exactly how many accounts out of 21. Final key value mapping shows frequency of following distribution. E.g 21:100, 20:50, 1:15 as an output means 100 accounts are following all 21 accounts of the list, 50 accounts are following 20 out of 21 accounts of list and 1:15 means that the mutual sum of all the followers who are following just one account of list is 15. Venn Diagram in 5.4 simplifies this. Sum of the shaded regions show sum of the mutually exclusive followers of all the accounts.

Using this algorithm, we figured out the number of followers that mapped onto the

**Algorithm 1** Followers Distribution

▷ Definitions
Let *user* be any vertex on a graph representation of a social network.
Let $A$ be a set of users and $N = |A|$. {In implementation, N = 100,000}
Let $\alpha$ be an ordered $N$-tuple that contains any permutation of $A$.
Let $frequency$ be any member of the set $\mathbb{Z}^+ \cup \{0\}$
Let $f : user \mapsto frequency$ map every unique user $u \in A$ to the number of occurrences of $u$ in $A$, and let $F \leftarrow \{f(u) : u \in A\}$.
Let $g : frequency \mapsto frequency$ map $f(u)$ to the number of occurences of $f(u)$ present in $F$, and let $G \leftarrow \{g(\epsilon) : \epsilon \in F\}$.

$H_0 \leftarrow$ HashTable $\{user \mapsto frequency\}$ {To compute $F$} every unique user $u \in A$ $H_0[u] = \emptyset$
$H_0[u] \leftarrow 1$
$H_0[u] \leftarrow H_0[u] + 1$

$H_1 \leftarrow$ HashTable $\{frequency \mapsto frequency\}$ {To compute G} every unique frequency $\epsilon \in F$ $H_1[\epsilon] = \emptyset$
$H_1[\epsilon] \leftarrow 1$
$H_1[\epsilon] \leftarrow H_1[\epsilon] + 1$

number of accounts. We were able to see the mutual convergence with all possible intersections, while keeping the output simple and easy to calculate. Another reason to use this algorithm is that we were interested in calculating the number of completely independent followers in addition to the number of totally common followers. The threshold of maxima and minima would have given us the knowledge of exclusiveness of accounts with each other. We took the top 21 accounts from our location, which included politicians, media persons, sportsmen, journalists and show business icons. Initially, we ran our experiment with the first 5000 followers and later upto 50,000 followers, and repeated this experiment thrice after 5 days on same timings. We wanted to ensure with an upper bound confidence, meaning that by 5 days, all our accounts would have gained at least 5000 new followers with no overlaps in the sample space. The results showed that the percentage of independent followers who followed just one particular account was always in the range of 11%-17%. The percentage of users who followed at least one other account was thus between 88%-83% (figures have been rounded off, actual being in decimals). The percentage of followers, following 21 accounts was also wavering between 18-35%.

To maintain fairness, we carried out the entire experiment by removing one figure from the 21 list and adding another renowned personality that was not part of the list. This was a control experiment. It was observed that the percentage of users following 21 accounts was 0 in the recent 5000 and 0.05% in 50000. So, our theory held a solid ground. However, the analysis of the entire data showed that the provision of followers was selective. Though the percentage of common âĂIJ21âĂİ was always higher than percentage of common âĂIJ20âĂİ, the mutual followers among 18 and 12 accounts had the highest percentage. But Since 21 accounts were constantly being fed followers by services, they remained at the top in terms of followers count, and hence, kept on coming in the recommendation for new accounts. Making the job easier for

the black market service providers

### 5.3 Experiment and Results

The experiment mentioned in section 1 was carried out again on these 21 accounts. The motive was to cross validate the theory involving malicious accounts, and to show that these accounts were not demonstrating human behavior. The accounts were checked for their followers growth and graphs were plotted. The results were almost identical for all the accounts and they had a strong cross correlation. In addition, we took the difference $\Delta d$ in their follower's count of 15 minutes. $\Delta d$ shows the change in their followers count within 15 minutes. When the entire vector of $\Delta d$ was plotted for all the accounts, they all showed almost exact pattern of rise or fall in followers. At some point in time, they all gained 10,000 followers within a span of 15 minutes and on another instant their followers count dropped by 4000 in the same window of time. Plotting the data from all accounts on a single graph would have been difficult to process visually, so we have plotted a chunk. However, the entire data set is available and can be obtained from the authors. Figure 4 displays the results.

## 5.4 FEATURE SET SORTING OF SUSPICIOUS ACCOUNTS

We picked one of the suspicious accounts on Twitter and did a statistical and nomenclature study on the followers of that account to build upon the previous work mentioned in Section 5.3. We then classified them based on those features set to determine the number of legitimate accounts and fake accounts. The account that we will mention in this paper is a verified account of a politician with 2.2 million followers.

### 5.4 Analysis and Results

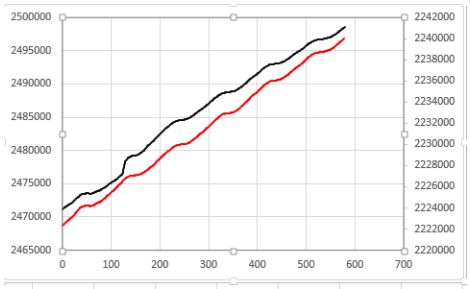
Our analysis on the target account showed that the out of 2.2 million followers 79.53 %(1.74 million) accounts had less than 10 tweets. 35%(0.77 million) accounts had 0 tweets. 92.93%(2.04 million) accounts had default profiles. 36% (0.792 million) accounts had default profile picture. 64% accounts had less than 10 followers. 48% accounts did not favorite any tweet ever. 80.4% accounts had no description or bio of their account. 6.5% accounts were being managed by third party apps like cloudhopper, tapbots, sprinkler, hootsuit,paper.li etc. We found around 283 unique third party apps which were controlling about 143,000 accounts. Some of the third party apps did not have ant web address or cyber presence. These are usually self created apps only managed by a set of users for particular intent.

### 5.4 Nomenclature Analysis

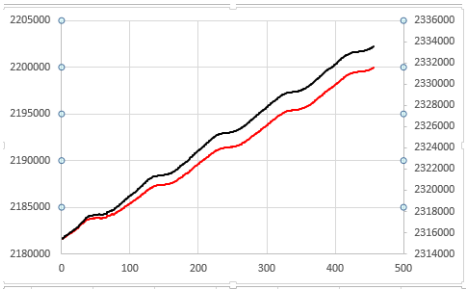
After primary classification, next we did nomenclature analysis of all the accounts. We found many peculiar things. It is likely to assume that there can be people of same name in cyber world. However if there are thousands of accounts with the name,

profile picture and description as that of former president **"Barack Obama**" then they are fake accounts. In our target accountâĂŹs followers, 39,060 accounts were of the same name and description as the the target account. This is to be noted that if these names are searched on Twitter, then only 24 accounts of that name appear. But results obtained from data fetched through API had 39,060 accounts of the same name. 21,071 accounts were of the same name as another famous politician of the same group. Thousands of accounts were bearing the same name and Bio as of famous celebrities, politicians and sports personalities. 19,122 accounts had the name âĂIJYoutube. 43,083 accounts had digits in their name. 7,822 accounts had the name âĂIJExpress News " and 4,892 accounts had the name âĂIJGeo News.âĂİ There were so many other amusing results. Mentioning all of them is beyond the scope of this paper.

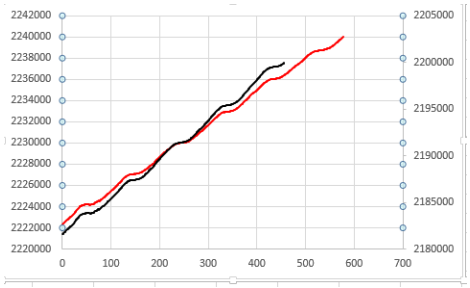
The key fact to understand here is that there are thousands of accounts which are counted in the followers of a profile. When these accounts are searched in the search box of Twitter, they do not come up. This shows that there is an entire community of fake followers or impersonators that is on Twitter. It does normal activities like following people and doing Retweets but is never shown on the main interface of Twitter. This begs the question that if Twitter is aware of presennce of these accounts then they should be removed. If Twitter classifies them as normal accounts, then they should be shown in the main search box.



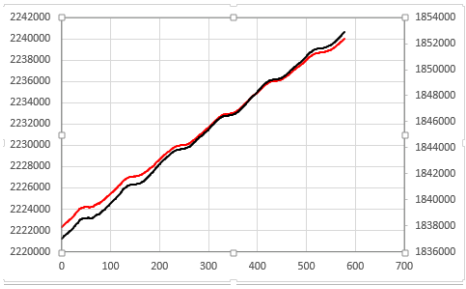

(a)



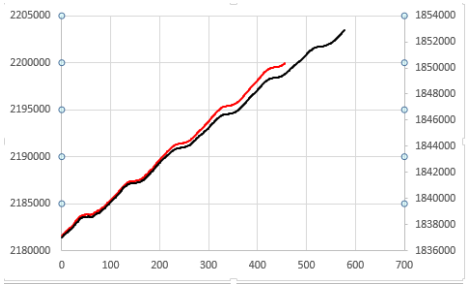

(b)



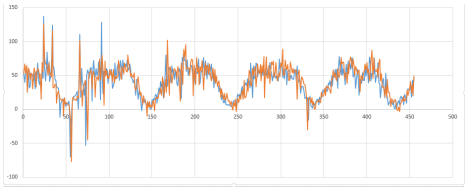

(c)



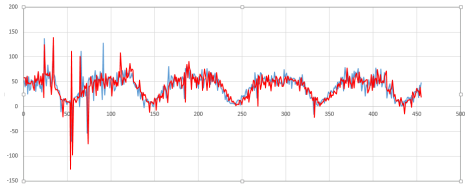

(d)



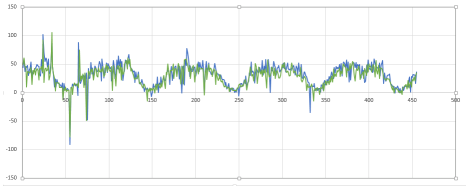

(e)



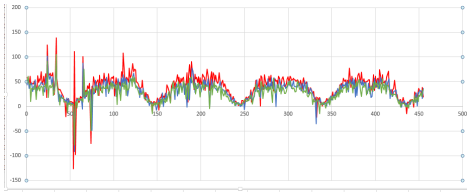
(f)



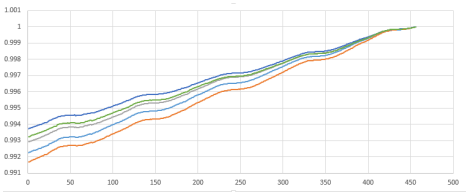
(g)



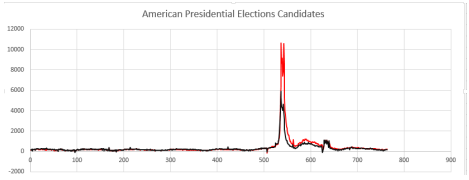
(h)



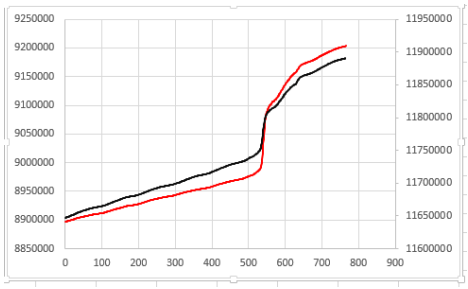
(i)



(j)




(k)




(l)

Figure 5.5: Top 21 Accounts

# Chapter 6

# Relevant Case Study

## 6.1 Case Study of 2014

In this chapter, we will present one case study that we did regarding political influence of Twitter in political scope of Pakistan. It is not an aspect of our research. It is merely a peculiar thing that we found on Twitter while doing our research.

Due to political unrest in August 2014, there was a massive outbreak on social media. People felt compelling need to be politically correct. Headlines of major news channels were swarmed with tweets and views. Social media started to influence mainstream media. It was realized by political parties to have a concrete presence on social media. In the attempt to gain more following on social media, they subscribed to the underground markets to buy fake followers. The intention was to demonstrate the muscle power on social media and show that educated class of the country had liking for them. We picked up one case study on Twitter and probed into the issue. The results were not very surprising. We did discover some unusual activities and unlikely boost in the followers count of our target account. As we indexed them with timeline of political events, the picture became very lucid. We show our experiment and result in Section 3.

### 6.1 August 2014's Political Scope

In August 2014, members and supporters of the opposition party in Pakistan started an agitation on the roads. They claimed that general elections of 2013 were heavily rigged. They demanded the sitting party in the Government to resign. There was a massive unrest in the country. The entire digital and print media covered the event. Social media was not immune to the effects of agitation. People heavily debated and argued about the ongoing political situation. They key thing was that the opposition party had a massive support on Twitter. Its leaders had a huge fan following on social media. They used it to voice their agendas and interact with their supporters. Very soon TV channels and newspapers started giving coverage to their Tweets and interactions. Still there are many newspapers who have a separate section in the paper where they print Tweets of politicians.

This did not go unnoticed by the sitting Government. So they also started to show their muscle on social media. Suddenly their followers count began to rise steeply. It was however peculiar that over the years these accounts did not have a substantial following but suddenly they started to appear very popular. Many people claimed that their followers were fake. Mostly because the new followers which were created never tweeted anything and never interacted with anyone. They appeared to be stillborn accounts. Due to an increased noise about legitimacy of their followers, we decided to do research in this field. We picked up a verified account of one of the leaders of sitting Government and carried out our experiment.

### 6.1 Data Gathering

We used two of the Twitter's REST API's to capture 2.2 million followers of our target account.

1. **GET followers/ids**
2. **GET statuses/user_timeline**

**GET followers/ids** returned the Ids of target account's followers. It returned a set of 5000 Ids in one get request. These Ids were sequentially passed to **GET statuses/user_timeline** from which we got all the relevant information of every account. We parsed the results and stored them in a file.

### 6.1 Experiment on Twitter

In the next step we sorted all the accounts with their creation dates. The objective was to see the distribution of followers with respect to their creation date. This would also give the idea of when that account started following our target. We used **'gnuplot'** to plot the followers with respect to their creation date. Figure below shows the output.

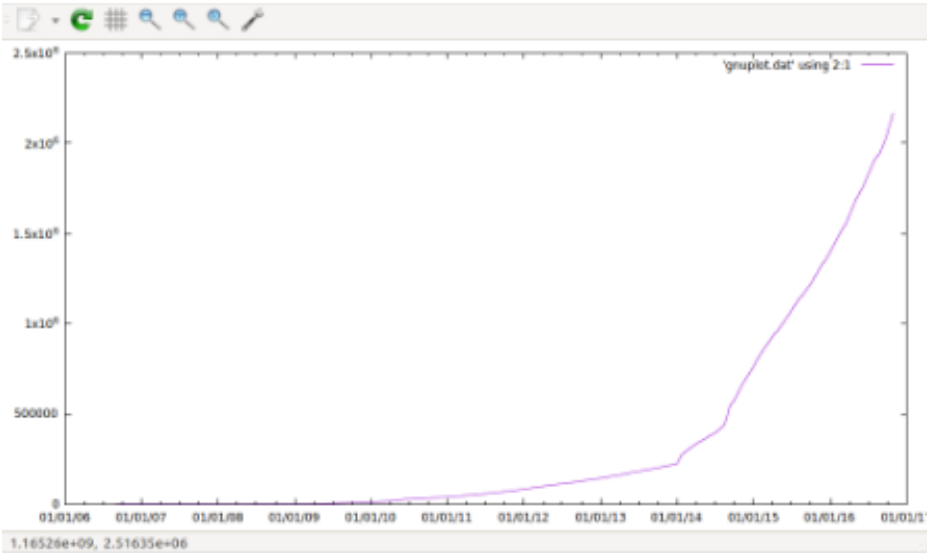



Figure 6.1: Growth of Target Account's Followers

The figure shows the creation date of accounts which followed the target account. The results can be interpreted easily. Before, August 2014, there were less than 25 thousand followers of target account. During August 2014, suddenly thousands of accounts were being created and they followed the target account. Within 6 months they had doubled and within next three months they had increased three times. So it was self evident that bots were being created or purchased and they were made to follow this account.

As mentioned earlier, people who claimed that the following of this account to be fake also accused that its followers were *"Stillborn."* In other words, they did not do any activity after their creation. They had no tweets and they did not interact with anyone. On the footprints of our first experiment we carried out our second experiment to test the claim. We sampled all the followers with *'0 status count.'* We arranged them with their creation dates as well. So the *'zero tweeters'* were sampled with time and the hypothesis was that their graph should have some common features

with the previous graph. If previously assembled followers followed a pattern, then provided that mostly them were bots, then *'zero tweeters'* should follow a similar trend in their creation dates. Figure 10 shows the result we obtained after plotting them.

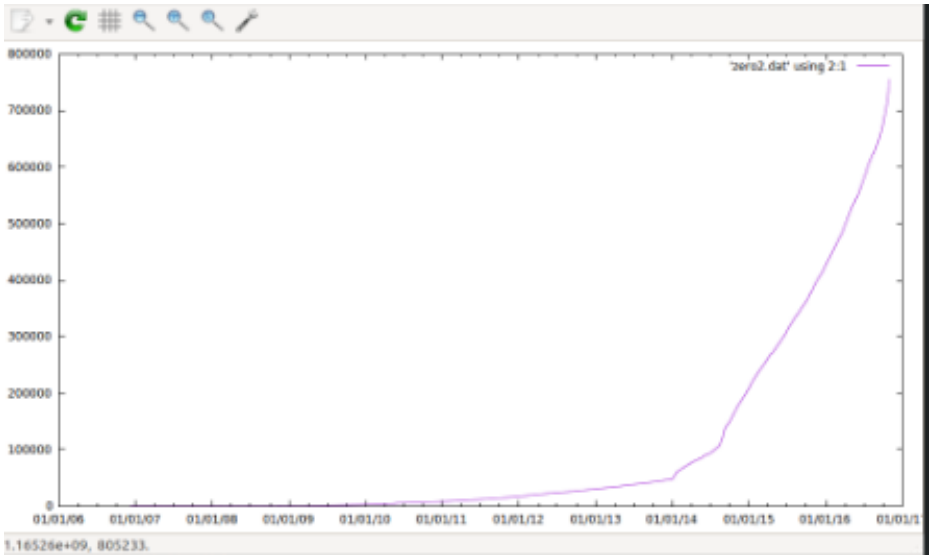



Figure 6.2: Distribution of Zero Tweeters

The second graph appeared to be almost a replica of the first graph. Thus adding weight to the argument that most of the followers gained by the target account after August 2014 were bots. These were not genuine organic followers. They were instead purchased to render fake popularity. In pursuit to compete with the opposition and to make a space in the social media, this account engaged into a malicious scheme. It is not surprising that majority of the social media users in Pakistan are not tech-savvy. Hence it is easy to coax them with fake news and fame. In our future work, we are using sentiment analysis to determine cluster of real users who persistently endorse or oppose certain narratives. The motive is to demonstrate relationship between people who hold similar views and find patterns in their clusters.

# Chapter 7

# Future Work and Conclusion

## 7.1 Third Party Applications and Cyber Crime

In our future work, we want to dig deep into the working and operations of third party applications. Once a user subscribes to a third party app a number of privileges are handed over to that app. Those privileges might include posting Tweets, sending Direct Messages or Retweeting other Tweets. Based upon the agreement that the author of the application has with Twitter, the application operates within the boundaries of that agreement. There are several fancy third party apps which use data to derive interesting analysis. Some of these apps like *TopFiveFollowers* tell the users about their top followers. Some other applications tell the user's ranking on Twitter or estimated worth of the account.

Once a third party is given the rights to post Tweets and update status, this has certain caveats too. If the author of the app is a malicious user, he can post unwanted things on behalf of actual user. In Pakistan, recently a cyber crime bill has been passed which imposes severe punishments on users who post blasphemous content on Twitter or Tweet against the interests of State. It is worth noticing that there are thousands of third party apps which have privileges of millions of users. If some of them are compromised or new apps are created on similar pattern and with malicious intent, then naive users are severely compromised. We want to carry out an experiment in which we highlight this issue and bring this to mainstream knowledge. There are several ways of identity impersonations on Online Social Networks and they can frame wrong people for cyber offence. This is the first objective in the work to come.

## 7.2 Solving the issue of Fake Followers

To solve the issue of fake followers, Twitter should employ following techniques.

- Remove all the **Skip** features in the sign-up phase.
- First verify the email and then make an account over that email.
- Solve the issue of ghost email IDs and permutations of email with **.** and **+**
- Remove all the accounts which are created over same email IDs with permutations
- Either show all the accounts with same names in the search result or remove them
- Add evaluation of third party applications by users on monthly basis
- Increase the layers of security in account creation and verification to stop automated scripts from creating such accounts.
- Detect the uniform growth of **Daily Followers** and flag them if found automated.
- Identify the lockstep Retweeting through third party applications and revoke the access of those third parties.
- Ensure an overall re-verification of all the accounts and users.

## 7.3 Conclusion

Twitter is a powerful social media tool with a massive impact over people. Recently it has been criticized for the volume of fake users and malicious content. Twitter has also been used by people to propagate their narratives via organic users and bots. Our work was to primarily outline the existence of collusion networks and come up with ways to spot automated and organic networks. We used *CopyCatch* algorithm to detect the lockstep behavior of human activity and Retweet source to find out bots. We also discovered beneficiary networks comprising of a number of collusion networks. This led to the disclosure of popular sentiments being engineered and the factions on Twitter which were highly involved in these activities. Another aspect of our work was to figure out the newly growing issue of "**Daily Followers**". We presented our strategy of catching and sorting these followers. We have identified the features of Twitter which allow black markets to create bulk followers and we suggest ways to overcome them. The overall effect of fake popularity, daily followers, collusion networks is engineered popularity on social media. This creates an effect of echo chamber for naive users who are surrounded by such collusion networks and there is not much awareness in the masses regarding this. Our thesis points out these caveats of social networks and we use computer science knowledge to establish a case for the study.

# References


[1] Carlo De Micheli and Andrea Stroppa, 2013. âĂIJTwitter and the underground market,âĂİ *11th Nexa Lunch Seminar* (Turin, 22 May)

[2] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummad *Measuring user influence on twitter: The million follower fallacy, 2010*

[3] Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. "What Is Twitter, a Social Network or a News Media?" *Proceedings of the 19th International Conference on World Wide Web - WWW '10 (2010):* n. pag. Web.11.

[4] B. A. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under the microscope. arXiv:0812.1045v1, Dec 2008.

[5] J. Ratkiewicz, M. D. Conover, M. Meiss, B. GoncÂÿalves, A. Flammini, F. Menczer. *Detecting and Tracking Political Abuse in Social Media*. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

[6] E. Ferrara, O. Varol, F. Menczer, A. Flammini. *The Rise of Social Bots*. Communications of the ACM, 2016.

[7] Beutel, Alex, Wanhong Xu, Venkatesan Guruswami, Christopher Palow, and Christos Faloutsos. "CopyCatch." *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 (2013)*: n. pag. Web.

[8] P. Dickerson, V. Kagan, V. S. Subrahmanian. *Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots?*. vol. 00, no. , pp. 620-627, 2014, doi:10.1109/ASONAM.2014.6921650

[9] Wang, A.H., 2010a. Detecting spam bots in online social networking sites: a machine learning approach Data and Applications Security and Privacy XXIV (pp. 335âĂŞ 342): Springer.

[10] Baik, Bok, Qing Cao, Sunhwa Choi, and Jin-Mo Kim. "Local Twitter Activity and Stock Returns." SSRN Electronic Journal (n.d.): n. pag. Web.

[11] Lee, Seunghun, Sungcheon Lee, and Hyun-Chul Kim. "Collecting Twitter Data using PlanetLab." (2014): n. pag. Web.

[12] Yang, C., Harkreader, R., Gu, G., 2013. Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. IEEE Trans. Inf. Forensics Security 8 (8), 1280âĂŞ1293

[13] SadBotTrue. N.p., n.d. Web. *https://www.sadbottrue.com*

[14] Adewole, Kayode Sakariyah, Nor Badrul Anuar, Amirrudin Kamsin, Kasturi Dewi Varathan, and Syed Abdul Razak. *"Malicious accounts: Dark of the social networks."* Journal of Network and Computer Applications 79 (2017): 41-67. Web.

[15] Funk, Tom. "Advertising and Promotion." Advanced Social Media Marketing (2012): 65-74. Web.

[16] "The Twitter Rules — Twitter Help Center." Twitter. Twitter, n.d. Web.

[17] Akoglu, L., Tong, H., Koutra, D., 2015. Graph based anomaly detection and description: a survey. Data Min. Knowl. Discov. 29 (3), 626âĂŞ688.

[18] Bhat, S. Y., Abulaish, M.2013. Community-based features for identifying spammers in online social networks. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[19] Harsule, S.R., Nighot, M.K., 2016. N-Gram Classifier System to Filter Spam Messages from OSN User Wall Innovations in Computer Science and Engineering. Springer, 21âĂŞ28

[20] "Latent Dirichlet allocation." Wikipedia. Wikimedia Foundation, 17 May 2017. Web. 05 June 2017.

[21] "The Most Reliable Place to Buy Twitter Daily Followers." Buyrealmarketing. N.p., n.d. Web. 12 Mar. 2017.

[22] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M., 2015. Fame for sale: efficient detection of fake Twitter followers. Decis. Support Syst. 80, 56âĂŞ71.

[23] "Topic: Twitter." *Www.statista.com* N.p., 31 Oct. 2016. Web.

[24] The Open Graph Viz Platform. (n.d.). Retrieved March 14, 2017, from https://gephi.org/