Analyzing Spammers' Social Networks for Fun and Profit

A Case Study of Cyber Criminal Ecosystem on Twitter

Chao Yang
Texas A&M University
College Station, Texas 77840
yangchao@cse.tamu.edu

Robert Harkreader
Texas A&M University
College Station, Texas 77840
bharkreader@cse.tamu.edu

Jialong Zhang
Texas A&M University
College Station, Texas 77840
jialong@cse.tamu.edu

Seungwon Shin
Texas A&M University
College Station, Texas 77840
swshin@cse.tamu.edu

Guofei Gu
Texas A&M University
College Station, Texas 77840
guofei@cse.tamu.edu

ABSTRACT

In this paper, we perform an empirical analysis of the **cyber criminal ecosystem** on Twitter. Essentially, through analyzing **inner social relationships** in the *criminal account community*, we find that criminal accounts tend to be socially connected, forming a small-world network. We also find that criminal hubs, sitting in the center of the social graph, are more inclined to follow criminal accounts. Through analyzing **outer social relationships** between criminal accounts and their social friends outside the criminal account community, we reveal three categories of accounts that have close friendships with criminal accounts. Through these analyses, we provide a novel and effective criminal account inference algorithm by exploiting criminal accounts' social relationships and semantic coordinations.

Categories and Subject Descriptors

k.6.5 [**Computing Milieux**]: Security and Protection; J.4 [**Computer Applications**]: Social and Behavioral Sciences

Keywords

Spammer, Online Social Network, Ecosystem

1. INTRODUCTION

Cyber criminals have utilized Twitter as a new platform to conduct their malicious behavior including sending spam and phishing scams [12], spreading malware [9, 5], hosting botnet command and control (C&C) channels [10], and launching other underground illicit activities. In March 2010, cyber criminals exploited Twitter to spread malware using festive-themed messages [9]. In September 2010, thousands of Twitter users including the wife of former British Prime Minister and White House Press Secretary were compromised by Twitter cyber criminals [12].

While most existing approaches [15, 26, 14, 32, 35] focus on detecting Twitter criminal accounts individually, we still understand far less about the properties of those criminal accounts' *social relationships* on Twitter. Yet, it is these very relationships that may be utilized by criminal accounts to increase their influence or to avoid detection and suspension.

WWW 2012, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1229-5/12/04.

Specifically, since Twitter users can automatically obtain their following accounts' updates, criminal accounts' social relationships can aid them in increasing the visibility of their malicious content – thus in obtaining more victims. In addition, by gaining more followers, Twitter criminal accounts can evade existing detection approaches such as "Twitter Rules" and break through Twitter's "Follow Limit Policy"[1], while maintaining their high visibility. Particularly, according to Twitter Rules [8], "a Twitter account can be considered to be spamming, and thus be suspended by Twitter, if it has a small number of followers compared to the amount of accounts that it follows."

However, we lack basic insights into the characteristics of criminal accounts' social relationships. How do criminal accounts socially connect with each other on Twitter? What is the topological structure of social relationships among those criminal accounts? Due to the fact that legitimate accounts normally do not like to follow criminal accounts, what are the main characteristics of criminal accounts' followers? Can we exploit these miscreants' tactics to build effective defense strategies against cyber criminals? The desire of addressing these questions empirically – and thus obtaining insights for defending against Twitter criminal accounts – forms the core motivation of this work.

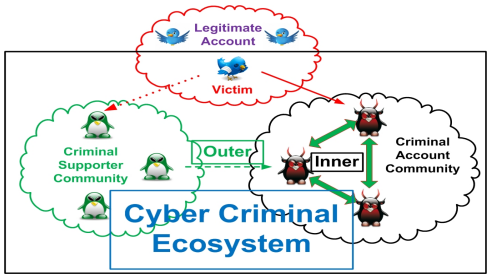


Figure 1: Structure of the cyber criminal ecosystem.

In this paper, we empirically analyze the **cyber criminal ecosystem** on Twitter, containing *criminal account community* composed of criminal accounts, and *criminal supporter community* composed of those accounts outside

[1]According to this policy, once an account has followed 2,000 users, the number of additional accounts it can follow is limited to its follower number [13].

the criminal account community who have close friendships (following relationships) with criminal accounts, defined in our work as criminal supporters (See Figure 1). Specifically, we analyze **inner social relationships** in the criminal account community to reveal insights on how criminal accounts socially connect with each other. Meanwhile, we analyze **outer social relationships** between criminal accounts and their criminal supporters to reveal the characteristics of those accounts who have close friendships with criminal accounts. We also aim at finding possible reasons why criminal supporters outside the criminal community become criminal accounts' followers. Essentially, these supporters aid criminal accounts in avoiding detection by increasing criminal accounts' followers, and in preying on more victims due to the "social-intercourse" nature of Twitter (Twitter users may visit their friends' friends' profiles). Through these analyses, we aim at understanding how criminal accounts mix into the whole Twitters space, and presenting new defense insights to effectively catch Twitter criminal accounts.

We conduct our empirical analysis based on a sample dataset containing around half million Twitter accounts with around 14 million tweets and 6 million URLs. After building a sample criminal account community composed of 2,060 identified spammer accounts in that dataset, we analyze its inner relationships by building and analyzing the social relationship graph. To analyze outer relationships, we propose a *Malicious Relevance Score Propagation Algorithm (Mr.SPA)* to extract criminal supporters. We then observe typical characteristics of three categories of supporters and provide possible reasons why these supporters have close friendships with criminal accounts. Finally, we design a *Criminal account Inference Algorithm (CIA)*, to infer unknown criminal Twitter accounts by starting from a seed set of known criminal ones and exploiting the properties of their social relationships and semantic coordinations with other criminal accounts.

In summary, the main contributions of our study are:

- We present the first in-depth case study of analyzing inner social relationships of criminal accounts. We have two main findings: (i) criminal accounts tend to be socially connected, forming a *small-world* network; (ii) compared with criminal leaves, criminal hubs are more inclined to follow criminal accounts.
- We propose a new algorithm *Mr.SPA* and have extracted 5,924 criminal supporters who have close friendships with criminal accounts. We also investigate the characteristics of three representative categories of criminal supporters. For example, we find that a representative category of such supporters, which we term as *social butterflies*, easily follow back any random account who initially follows them (in our test, about 48% of them do so within 48 hours), while in reality very few (less than 2%) normal accounts would do this. *This implies that by initializing social relationships with these kind of accounts, criminal accounts can easily mix into Twitter.*
- We design a new algorithm CIA to infer more criminal accounts based on a small known seed set, by simply analyzing the social relationships and semantic coordinations among accounts. *Using CIA, we can infer over **20** times more criminal accounts than that of using a random selection strategy.*

2. RESEARCH GOAL AND DATASET

2.1 Research Goal

Our research goal is to provide the first empirical analysis on *how criminal accounts mix into and survive in the whole Twitter space*. Specifically, we target on those criminal accounts as defined by Twitter Rules [8], who mainly post malicious URLs linking to malicious content with an intention to compromise users' computers or privacy. Through analyzing inner social relationships in the criminal community composed of criminal accounts (in Section 3), we aim at answering the following questions. What is the structure of criminal accounts' network? What are possible factors and inherent reasons leading to that structure? Are there any different social roles for different types of criminal accounts? Through analyzing outer social relationships (in Section 4), we aim at answering the following questions. What are typical characteristics of the accounts outside the criminal community that tend to follow criminal accounts? What are possible reasons that these accounts have close friendships with criminal accounts? Then, through exploiting criminal accounts' social relationships, we design an inference algorithm to catch more criminal accounts (in Section 5).

2.2 Dataset

To achieve our research goals, we analyze the dataset from our previous Twitter spam account detection study [35], which is crawled by tapping into Twitter's Streaming API [11] from April 2010 to July 2010. The dataset contains 485,721 Twitter accounts with 14,401,157 tweets and 5,805,351 URLs. Due to the large amount of shortening URLs on Twitter, for each URL in every tweet, the dataset records its final destination URL through following the URL redirection chain.

To analyze criminal accounts, we also use the results from that previous study [35], which outputs 10,004 malicious affected accounts posting malicious URLs. Of those malicious affected accounts, 2,060 accounts are finally identified as spammer accounts. The URLs are labeled as malicious by using the widely-used URL blacklist Google Safe Browsing (GSB) [4] and a high-interaction client honeypot, implemented using Capture-HPC [3]. We clearly acknowledge and discuss the limitations of our analyzed dataset in Section 7.

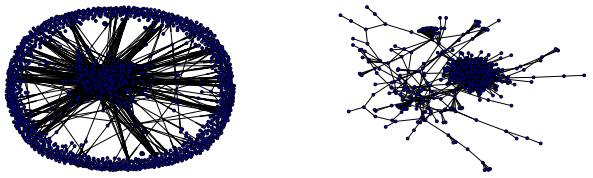
Based on this dataset, we build and analyze a sample criminal account community, which is composed of those 2,060 identified spam accounts.

3. INNER SOCIAL RELATIONSHIPS

In this section, we empirically analyze inner social relationships in our sample criminal account community by visualizing its relationship graph and revealing its relationship characteristics.

3.1 Visualizing Relationship Graph

If we view each criminal account as a node v and each follow relationship as a directed edge e, we can view inner social relationships in the criminal account community on Twitter as a directed graph, named as the criminal relationship graph $G = (V, E)$. In our dataset, the criminal relationship graph consists of 2,060 nodes and 9,868 directed edges (see Figure 2(a)). By further breaking down the graph, we can obtain 8 weakly connected components containing at least three nodes and 521 isolated nodes. (Since we can partially crawl the whole Twitters space and utilize a relatively strict way of identifying criminal accounts, the number of isolated



(a) Relationship graph (b) Connected component

Figure 2: Criminal relationship graph. Each "*dot*" represents a criminal account and each "*line*" connects a pair of following and follower criminal account. The more relationships an account has, the more central it is positioned in the graph.

accounts may be somewhat overestimated.) The giant connected component contains 954 nodes (see Figure 2(b)).

3.2 Revealing Relationship Characteristics

After visualizing our sample criminal relationship graph, we analyze this graph through utilizing graph theoretical knowledge, and observe the following two main findings.

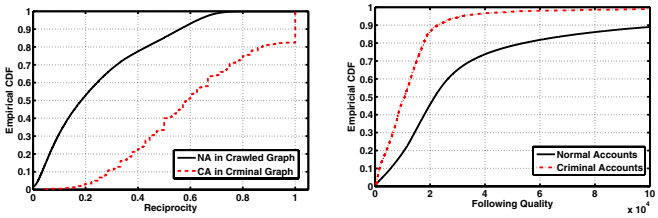
Finding 1: Criminal accounts tend to be socially connected, forming a small-world network. From Figure 2(a), we can observe that criminal accounts tend to socially connect with each other. To quantitatively validate this finding, we measure three graph metrics: graph density, reciprocity, and average shortest path length.

Graph density is the proportion of the number of edges in a graph to the maximal number of edges, which can be computed as $\frac{|E|}{|V|\cdot(|V|-1)}$. This metric measures how closely a graph is to be a complete graph. A higher value implies that the graph is denser. After calculating the graph density for both our sample criminal relationship and a public entire Twitter snapshot[25] containing 41.7 million users and 1.47 billion edges, we find that the graph density of our sample criminal relationship graph, which is 2.33×10^{-3}, is much higher than that of the Twitter snapshot, which is 8.45×10^{-7}. This shows that the criminals have closer relationship than regular Twitter users.

Reciprocity is represented by the number of bi-directional links[2] to the number of outlinks. We find that criminal accounts have stronger reciprocal social relationships than legitimate accounts. For example, around 95% criminals accounts have the reciprocity higher than 0.2 in the criminal graph, while only around 55% normal accounts in our crawled graph (containing around $500K$ nodes) [35] have such values (See Figure 3(a)). Furthermore, around 20% of criminal accounts' values of reciprocity in the criminal graph are nearly 1.0, i.e., *all* criminal accounts followed by these 20% of criminal accounts also follow them back. This observation clearly implies that criminal accounts have stronger social relationships in the criminal account community.

Average Shortest Path Length is defined as the average number of steps along the shortest paths for all possible pairs of graph nodes. It can be used to measure the efficiency of information flow on a graph. Compared with the average path length of a sample data set with 3,000 legitimate Twitter accounts [25], which is 4.12, the average shortest path length of the criminal relationship graph is even smaller, which is 2.60. This implies that the criminal account community is also a small-world network. As an important property, a small-world network contains a giant connected component, which can be verified in Figure 2(b).

From the above analysis, we can find that criminal accounts have strong social connections with each other. Then, the next question we try to answer is: *what are the main factors (criminal accounts' actions) leading to that structure?*



(a) Reciprocity (b) Following Quality

Figure 3: The comparison of the criminal accounts (CA) and normal accounts (NA).

Possible Factor 1: *Criminal accounts tend to follow many other accounts without considering those accounts' quality much, making themselves to connect to other criminal accounts.* The observation that criminal accounts tend to follow many other accounts has been widely analyzed and utilized to build detection schemes in existing work [26, 14, 33, 32, 35]. Our work focuses on analyzing the quality of criminal accounts' followings. Specifically, we use an account's follower number to reflect its quality, i.e., intuitively, an account with more followers has higher quality. (We acknowledge that an account's follower number may not accurately measure its quality in all cases. However, it is still a practical and well-accepted metric to laterally reflect an account's fame and reputation [2].) Then to measure the quality of an account's following accounts, we use a metric, named "*following quality*", which is the average follower number of an account's all following accounts. In this way, a higher *following quality* of an account implies that this account tends to follow those accounts with more followers. We can find that compared with normal accounts, criminal accounts tend to follow accounts with fewer followers (See Figure 3(b)). Around 85% of criminal accounts have the following quality lower than 20,000, while only around 45% of normal accounts have such a value. This observation validates that criminal accounts' actions of indiscriminately following others lead them to connect with low quality accounts, and hence connect with other criminal accounts.

Possible Factor 2: *Criminal accounts, belonging to the same criminal organizations, may be artificially/intentionally connected with each other.* To validate this possible factor, similar to [22], we first group criminal accounts into different criminal campaigns/clusters (possibly denoting different organizations) by clustering them with their posted malicious URLs. Each criminal campaign contains the criminal accounts that post the same unique malicious URLs. In this way, we obtain 17 criminal campaigns and each of them has at least 3 nodes. Then, for each campaign, we draw its criminal relationship graph using a similar method to the one mentioned in Section 3.1 and calculate its number of edges. The sum of those numbers of edges is 8,667, which is around 87.8% of all edges in the whole criminal commu-

[2]There is a bi-directional link between two nodes, if they reciprocally link to each other.

nity. The observation that many social edges are built within criminal campaigns (rather than across different campaigns) implies strong social connections within criminal campaigns.

Although it is difficult to accurately trace how these connections are generated, this observation still reflects the high probability that criminal accounts in the same criminal organization are artificially/intentionally connected.

In fact, no matter whether these connections are built using random selection or intentional construction, criminal accounts could benefit from such strong social connections in the criminal community. Essentially, this structure provides "support" (followers) to criminal accounts, which are very important for criminal accounts to either break the Following Limits Policy or evade detection features that are built based on the metric of follower number.

As seen in Figure 2(a), some nodes in the middle of the graph, termed as *criminal hub* in our work, have a bunch of social relationships with the nodes in the periphery, termed as *criminal leaf*. *Are there any differences between these two types of criminal accounts?* This motives us to study special characteristics of those criminal hubs. Before analyzing those hubs, we extract criminal hubs by calculating hub scores of criminal accounts in terms of their positions in the criminal relationship graph by utilizing the HITS algorithm [23]. Particularly, for each vertex i in the graph, we can use Eq.(1) and (2) to compute its hub score H_i^t and authority score A_i^t in the t-th iteration. When the computation converges within several iterations, we can obtain this vertex's final hub score H_i.

$$H_i^t = \begin{cases} 1, & if\ t = 0 \\ \sum_{(i,j)\in E} A_j^{t-1}, & if\ t > 0 \end{cases} \quad (1)$$

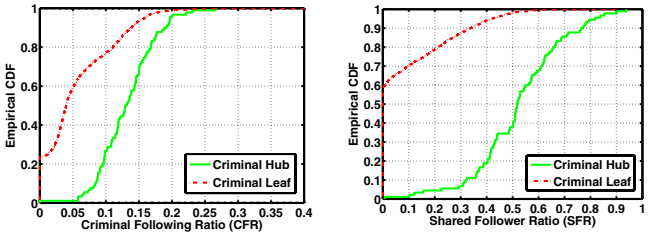
$$A_i^t = \begin{cases} 1, & if\ t = 0 \\ \sum_{(j,i)\in E} H_j^{t-1}, & if\ t > 0 \end{cases} \quad (2)$$

According to this algorithm, a higher hub score of an account implies that it follows many accounts with high follower numbers. Thus, we extract 90 criminal hubs with higher scores and 1,970 criminal leaves with lower scores by using k-means algorithm and setting $k = 2$.

Finding 2: Compared with criminal leaves, criminal hubs are more inclined to follow criminal accounts. To validate this finding, we examine whether criminal hubs' followings are more likely to be criminal accounts. For better description, we term a criminal account's following account as a "criminal-following", if this following account is also a criminal. Then, we design a metric, named Criminal Following Ratio (CFR), which is the ratio of the number of an account's criminal-followings to its total following number. A higher CFR of an account implies that this account is more inclined to follow criminal accounts. From Figure 4(a), we can find that criminal hubs' CFRs are much higher than that of criminal leaves. Specifically, around 80% criminal hubs' CFRs are higher than 0.1, while only 20% of criminal leaves have such values. Also, almost no criminal hubs' CFRs are lower than 0.05, while around 60% of criminal leaves' CFRs are lower than 0.05. This observation validates that criminal hubs tend to follow more criminal accounts than leaves do. Similar to Finding 1, we next provide possible explanations to Finding 2.

Possible Explanation: *Criminal hubs tend to obtain followers more effectively by following other criminal accounts (to obtain their followers).* Although criminal accounts could obtain followers by randomly following any account and expecting it to follow back, this method is still not very effective, due to the low chance of successfully alluring legitimate accounts to follow back. However, through following criminal accounts, hubs can automatically acquire those criminal accounts' followers' information (Username or Account ID). Then, there is a better chance for criminal hubs to successfully allure other criminal accounts' followers to become their own followers. Particularly, these followers have been already proved to be more susceptible to follow criminal accounts, which many legitimate accounts may not choose to do. (For more details supporting this argument, refer to Section 4.) In this way, criminal hubs can obtain followers more effectively.

To validate this explanation, we examine whether criminal hubs' followers are highly shared with their criminal-followings. Specifically, we design a metric, named Shared Follower Ratio (SFR), which is the percentage of an account's followers, who also follows at least one of this account's criminal-followings. A high SFR of an account implies that most of this account's followers are also its criminal-followings' followers, i.e., this account tends to share common followers with its criminal-followings. We find that criminal hubs' SFRs are higher than criminal leaves'. Around 80% of criminal hubs' SFRs are higher than 0.4, while around 5% of criminal leaves have such values (see Figure 4(b)). This observation reflects that compared with criminal leaves, criminal hubs' followers share more followers with their criminal-followings. This indirectly implies that criminal hubs could obtain followers by knowing their criminal-followings' followers' information.



(a) Criminal Following Ratio (b) Shared Follower Ratio

Figure 4: The comparison between criminal hubs and criminal leaves.

From these two findings, we can roughly draw a picture on how criminal accounts could obtain followers on Twitter. *Similar to the Bee Community*, in the criminal account community, criminal leaves, *like bee workers, mainly focus on collecting pollen (randomly following other accounts to expect them to follow back)*; criminal hubs in the interior, *like bee queens, mainly focus on supporting bee workers and acquiring pollen from them (following leaves and acquiring their followers' information).*

4. OUTER SOCIAL RELATIONSHIPS

If criminal accounts mainly build inner social relationships within themselves, according to existing approaches such as Sybil Guard [37] and Sybil Infer [19] on detecting sybil nodes, criminal accounts can be easily detected. However, many Twitter criminal accounts have already utilize several tricks to obtain followers outside the criminal account community and mix well into the whole Twittersphere [7].

Those accounts outside the criminal community, who have close "follow relationships" with criminal accounts, essentially aid criminal accounts both in avoiding detection and in spreading malicious content [6]. In our work, we define them as *criminal supporters*.

However, currently we have little knowledge about the characteristics of these criminal supporters. Thus, in this section, we conduct the first analysis of outer social relationships between criminal accounts and their supporters. By extracting and characterizing criminal supporters, we reveal typical characteristics of these supporters and understand more on how criminal accounts could mix into the Twitter space.

4.1 Extracting Criminal Supporters

We first design a *Malicious Relevance Score Propagation Algorithm (Mr.SPA)* to extract criminal supports. Specifically, Mr.SPA will assign a malicious relevance score (MR score) to each Twitter account, measuring how closely this account follows criminal accounts. A higher MR score implies a closer "follow relationship" to criminal accounts. Then, we measure the MR score based on three heuristics: (1) the more criminal accounts that an account has followed, the higher score this account should inherit; (2) the further an account is away from a criminal account, the lower score the account should inherit; (3) the closer the support relationship between a Twitter account and a criminal account is, the higher score the account should inherit.

To formalize the above intuitions, we build a *malicious relevance graph* $G = (V, E)$ to model the support relationship. In this graph, we consider each Twitter account i in our dataset outside the criminal community as a node V_i. There is a directed edge e_{ij} from the node V_i to the node V_j, if the account i follows the account j. The weight W_{ij} of the edge e_{ij} is determined by the closeness of the relationship between i and j. We next introduce our malicious relevance score propagation algorithm, which contains two phases: initializing MR score, and propagating MR score.

MR Score Initialization: Before propagating MR score, we first assign an initial score M_i^0 to each node V_i. If we denote $C = \{C_i | C_i \text{ is a criminal account}\}$, then each criminal account $C_i \in C$ is assigned a non-zero score m_i[3]. For other accounts, the score is initialized to zero.

MR Score Propagation: To propagate a MR score M_i to each node V_i after the initialization phase, we make the following three score-assigning policies according to the above three heuristics:

Policy 1: MR Score Aggregation. An account's score should sum up all the scores inherited from the accounts it follows. As Figure 5(a) illustrates, when A follows both criminal accounts C_1 and C_2, the score of A is the sum of the malicious scores of C_1 and C_2.

Policy 2: MR Score Dampening. The amount of MR score that an account inherits from other accounts should be multiplied by a dampening factor of α according to their social distances, where $0 < \alpha < 1$. As Figure 5(b) illustrates, when A_1 is one hop away from a criminal account C, we assign it a dampening factor of α, where $0 < \alpha < 1$. When A_2 is two-hop away, A_2 will get a dampening factor of $\alpha \cdot \alpha = \alpha^2$.

Policy 3: MR Score Splitting. The amount of MR score that an account inherits from the accounts it follows should be multiplied by a relationship-closeness factor W_{ij}, which is the weight of the edge in our *malicious relevance graph*. Specifically, we use the number of followers of an account to reflect the closeness of the relationship between this account and its followers. (The intuition is that if an account has more followers, the closeness of the relationship between this account and each of its followers will become weaker.) As Figure 5(c) illustrates, if A_1 and A_2 have followed the same criminal account C, the relationship-closeness factor of each account to C is 0.5. Thus, according to this policy, the score of a node V_i can be computed as $M_i = W_{ij} \cdot M_j,\ if\ (i,j) \in E$.

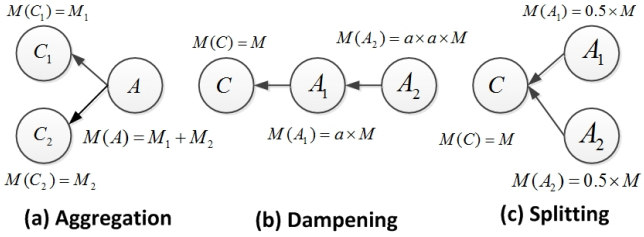


Figure 5: The policies of assigning MR scores.

Before presenting our mathematical model of propagating MR score, we first introduce some notations. Let n be the number of nodes in the *malicious relevance graph*. We use the indication function $I_{ij} = \{0, 1\}$ to indicate whether $(i,j) \in E$ (i.e., if $(i,j) \in E$, $I_{ij} = 1$; otherwise, $I_{ij} = 0$). If we use $numIndegree(j)$ to denote the number of the in-degree of the node j, then from *MR Score Splitting* policy, we can obtain that $W_{ij} = \frac{1}{numIndegree(j)}$. We use $\mathbf{I}$ to denote the column-vector normalized adjacency matrix of nodes (i.e., $\mathbf{I}_{ij} = I_{ij} \cdot W_{ij}$, if $numIndegree(j) \neq 0$; $\mathbf{I}_{ij} = \frac{1}{n}$, if $numIndegree(j) = 0$). Let $\overrightarrow{\mathbf{M}^0}$ be initial MR Score vector for all nodes and let $\overrightarrow{\mathbf{M}^t}$ be malicious score column vector for all nodes at the step t.

According to those three policies and our notations, at each step, for each node V_i, its simple MR score M_i can be computed using Eq.(3).

$$M_i = \alpha \cdot \sum_{j=1}^{n} I_{ij} \cdot W_{ij} \cdot M_j \tag{3}$$

In addition, with the consideration of each node's historical score record, at each step $t(t > 0)$, we add an initial score bias $(1-\alpha) \cdot M_i^0$ to its simple MR Score. (In our experiment, we set $\alpha = 0.85$, since it is widely used in the random-walk model.) Thus, we can compute the MR Score column-vector $\overrightarrow{\mathbf{M}^t}$ for all nodes at the step $t(t > 0)$ by Eq.(4).

$$\overrightarrow{\mathbf{M}^t} = \alpha \cdot \overrightarrow{\mathbf{I} \cdot \mathbf{M}^{t-1}} + (1-\alpha) \cdot \overrightarrow{\mathbf{M}^0}\ (t > 0) \tag{4}$$

When the score vector converges after several propagation steps, we can obtain final MR scores for all nodes. Once all MR scores have been calculated, a threshold is needed to determine which accounts have *sufficiently* close friend relationships to their criminal counterparts. To find an acceptable threshold, we first use x-means algorithm [29] to cluster accounts based on their MR scores. In this way, accounts with similar scores will be grouped together indicating they have similar follow relationships with criminal accounts. Then, we observe that most accounts have relatively small scores and are grouped into one single cluster. That is mainly because most accounts do not have very close follow relationships with criminal accounts. With this observation, we choose the highest score of the account in that

[3]In our preliminary experiment, we set $m_i = 1$.

cluster as the threshold. Then, we output 5,924 criminal supporters, whose MR scores are higher than the threshold.

4.2 Characterizing Criminal Supporters

After extracting criminal supporters, according to our empirical studies, we observe three representative categories of supporters (social butterflies, social promoters, and dummies) according to our defined thresholds. (Since we aim at showing preliminary and basic insights of criminal supporters' characteristics, the thresholds that are used to characterize them can be tunable according to how strictly to reflect their behavioral characteristics.)

Social Butterflies are those accounts that have extraordinarily large numbers of followers and followings. Like social butterflies in our real life, these accounts build a lot of social relationships with other accounts without discriminating those accounts' qualities. To qualitatively define social butterflies, we use 2,000 following as a threshold in terms of Twitter's *Follow Limit Policy* [13]. This number could be efficiently used to distinguish whether an account is socialized. In this way, we find 3,818 social butterflies.

We present our hypothesis that *the reason why social butterflies tend to have close friendships with criminals is mainly because most of them usually follow back the users who follow them without careful examinations.* Especially, some public software and services can help users automatically follow back other users, who have followed them [1]. In this way, these social butterflies might unintentionally follow back criminal accounts upon requests.

To validate this hypothesis, we first sign up 30 Twitter accounts without any tweets and any personal information. Then we use 10 accounts to follow 500 accounts (each account follows 50 accounts) that are randomly selected from those 3,818 butterflies. Meanwhile, we use another 10 accounts to follow another randomly selected 500 normal accounts without posting any malicious tweets, and the other 10 accounts to follow another randomly selected 500 identified criminal accounts. To minimize the influence generated by our experiment, we close our signed-up accounts after 48 hours. During this timespan, we find 47.8% of those butterflies follow back to our signed-up accounts, while only 1.8% of those normal accounts and 0.6% of those criminal accounts follow back. The fast speed in which these social butterfly accounts followed our accounts back validates our hypothesis that these accounts may automatically follow back any accounts that follow them. Such a low value for those criminal accounts validates that our identified criminal accounts are not social butterflies. And they usually will not follow back other accounts, since this behavior will not increase their follower numbers and influence. This experiment also shows that even though those Twitter accounts with many followers are usually popular and trustable, we cannot always trust their friends' quality.

Social Promoters are those Twitter accounts that have large following-follower ratios (the ratio of an account's following number to its follower number), larger following numbers and relatively high URL ratios. The owners of these accounts usually use Twitter to promote themselves or their business. We extract those social promoters whose URL ratios (the ratio of the number of URLs to the number of tweets) are higher than 0.1, and following numbers and following-follower ratios are both at the top 10-percentile of all accounts in our dataset. In this way, we can obtain 508 social promoters.

We make our hypothesis that *the reason why social promoters tend to have close friendships with criminal accounts is probably because most of them usually promote themselves or their business by actively following other accounts without considerations of those accounts' quality.* Thus, promoters may become criminal supporters by unintentionally following criminal accounts.

For this type of supporters, we use a heuristic method to validate our hypothesis. Since the goals of these promoters are promoting themselves or their business, they usually repeat posting URLs with the same domain names, which link to the webpages containing their promotion information. Thus, the purity of domain names in promoters' posted URLs are higher, leading a lower domain name entropy. With this intuition, to calculate domain name entropy for each social promoter, we extract each promoter's posted domain names in the final URLs, which are obtained through following URL redirection chains, due to the wide usage of shortening URLs on Twitter. Then, we compute its domain name entropy by using $-\sum_{i=1}^{N} p_i \ln p_i$, where N denotes the number of distinct domain names and p_i denotes the ratio of the occurrences of the i-th distinct domain name to the total number of domain names.

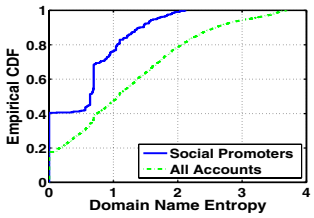


Figure 6: The entropy of the domain names.

Figure 7: A Case study for social promoters.

From Figure 6, we can find around 40% social promoters' domain name entropy are zero, which implies that all their URLs have the same domain names. Also, social promoters' domain name entropy are lower than that of other accounts. Specifically, around 80% social promoters have the domain name entropy lower than 1.0, whereas around 45% of all accounts in our dataset have such values. The observation heuristically validates our hypothesis that supporters tend to use Twitter to promote themselves by actively following other accounts, leading to close relationships with criminal accounts. One case study for a social promoter can be seen in Figure 7. The owner of this promoter mainly utilizes Twitter to promote an online book selling website.

Dummies are those Twitter accounts who post few tweets but have many followers. Since in Twitter, legitimate users tend to follow those accounts that share more useful information, it is relatively weird that these dummies has close relationships with criminal accounts while sharing little information, but they have many followers. In particular, we extract intriguing dummy accounts, who post fewer than

5 tweets[4] and whose follower numbers are at the top 10-percentile. In this way, we obtain 81 dummies.

We make our hypothesis that *the reason why dummies intend to have close friendship with criminals is mainly because most of them are controlled or utilized by cyber criminals.* To validate this, we analyze these dummy accounts several months after the data collection. Then, we find that 1 account has been suspended by Twitter, and 6 accounts do not exist any more (closed), and 36 accounts begin posting malware URLs labeled by Google Safe Browsing, and 8 accounts begin posting (verified) phishing URLs.

The observation that dummies begin posting malicious URLs shows that these dummies who have close relationships with cyber criminals are highly likely controlled by cyber criminals. A case study of one dummy account, who posted no tweets at the time when we crawled its profile but starts to post malicious tweets later, can be seen in Figure 8. This dummy account steals victims' email addresses through claiming to help people earn money. However, the dummy account sends email spam.

Figure 8: A case study for dummies.

Also, we find that unlike social butterflies and promoters, dummies are a special type of supporters extracted by Mr.SPA, since they initially do *not* post malicious URLs. However, they may later evolve to be criminal accounts. This discrepancy is mainly because our work provides a static view of the ecosystem. We note that *criminal accounts could be dynamically evolved from those dummies extracted by Mr.SPA*, and thus we do not argue whether dummies are supporters or criminal accounts.

Through analyzing outer social relationships between criminal accounts and their supporters, we can understand more on how criminal accounts could mix into the whole Twitter space by achieving criminal supporters. Also, *once we extract these supporters, we can warn legitimate users not to make friends with these supporters so as to avoid exposure to criminal accounts.*

5. INFERRING CRIMINAL ACCOUNTS

Considering the huge number of Twitter accounts, it is impractical to make in-depth checks on every account whether it is a criminal account at the same time. A lightweight inference algorithm, to guide to more suspicious accounts instead of scanning or analyzing all accounts given limited resources or time, is indeed needed. (In fact, similar existing work such as [38] mainly utilizes a ranking mechanism to predict potential threats.) As criminal accounts tend to be socially connected, a spontaneous and practical strategy is to first check those accounts that are connected with known criminal accounts by using Breadth First Search (BFS) algorithm. In this section, we propose a *Criminal accounts Inference Algorithm (CIA)* to infer more criminal accounts by exploiting criminal accounts' social relationships and semantic coordinations.

[4]None of these tweets contain URLs that are labeled as malicious by GSB or honey client.

5.1 Design of CIA

In brief, our Criminal account Inference Algorithm (CIA) propagates malicious scores from a seed set of known criminal accounts to their followers according to the closeness of *social relationships* and the strength of *semantic coordinations*. If an account accumulates sufficient malicious score, it is more likely to be a criminal account.

The intuition of CIA is based on the following two observations: (1) criminal accounts tend to be socially connected; (2) criminal accounts usually share similar topics (or keywords or URLs) to attract victims, thus having strong semantic coordinations among them. The first observation has been shown and discussed in Section 3. The second observation has also been widely analyzed in existing work such as [22, 21], which validates the existence of shared semantic topics among different criminal account campaigns.

In general, our CIA integrates the first observation by referring to *Mr.SPA* designed in Section 4 to quantify the closeness of social relationships. To integrate the second observation, we design a metric, *Semantic Similarity score (SS score)*, among each pair of accounts to quantify their semantic coordinations. To calculate SS score, we first extract a *Semantic Fingerprint Vector (SFV)* for each account, which essentially contains several representative terms in its tweets based on the TF-IDF algorithm [31], a widely used metric in the information retrieval community to measure the representativeness of terms. Then, SS score of each pair of accounts can be computed as the distance of their SFVs. (Due to the page limitation, we omit the technical details of extracting SFV and calculating SS score. Here, we just note that a higher SS score between two accounts implies that they have stronger semantic coordinations.)

With the above intuitions and notions, we then describe the design of CIA in details. To infer criminal accounts in a set of U Twitter accounts, we first start from a known seed set of M criminal accounts. Then, similar to *Mr.SPA*, we build a malicious relevance graph by using these $(M+U)$ accounts, denoted as $G = (V, E)$. In this graph, each account denotes a vertex in V and each follow relationship denotes a directed edge in E. Then, unlike *Mr.SPA*, we assign a weight for each edge $e_{ij} \in E$ (by using a semantic weight assignment function $WS(i,j)$) to reflect the semantic coordination between each pair of accounts. The basic intuition of designing this function is that if an account has higher SS scores (stronger semantic coordination) with its followings, it should inherit more malicious score from its followings. With this intuition, for each account j, we calculate every SS score between itself and each of its follower account i, denoted as SS_{ij}. Then, the weight $WS(i,j)$ of the edge e_{ij} can be calculated as $WS(i,j) = \frac{SS_{ij}}{\sum_{e_{kj} \in E} SS_{kj}}$.

Then, similar to *Mr.SPA*, for each criminal account, we assign a non-zero malicious score and propagate this score by using the semantic weight assignment function $WS(i,j)$ in Eq. (3). In this way, we can see that an account's malicious score can be proportionally distributed to its followers according to the closeness of social relationships and strength of semantic coordinations. When the score vector converges after several propagation steps, we infer those accounts with high malicious scores as criminal accounts.

5.2 Evaluation of CIA

We evaluate our Criminal account Inference Algorithm

(CIA) based on two different datasets – Dataset I and Dataset II. Dataset I refers to the one with around half million accounts from our previous study [35]. Dataset II contains another new crawled $30K$ accounts by starting from 10 newly identified criminal accounts and using breath-first search (BFS) strategy.

To evaluate the effectiveness of our CIA, similar to [38] that uses the number of hits in a top list, we use the number of correctly inferred criminal accounts and malicious affected accounts (denoted as CA and MA, respectively) in a top (ranked) list. (Even though these malicious affected accounts may not be criminal accounts, they still pollute Twitter with malicious URLs and create a risk for innocent users.) Thus, a higher number of CA and MA indicates that the algorithm is more effective to infer criminal accounts.

Note that as a lightweight *inference and ranking* algorithm aiming at magnifying suspicious accounts from a small seed set, we do *not* position CIA as a full *detection* algorithm. Thus, we adopt similar metrics to "Hit Count" used in [38] to measure CIA's effectiveness rather than using false positive and false negative rate. However, CIA could definitely be incorporated into some actual criminal account detection system by combining with other detection features.

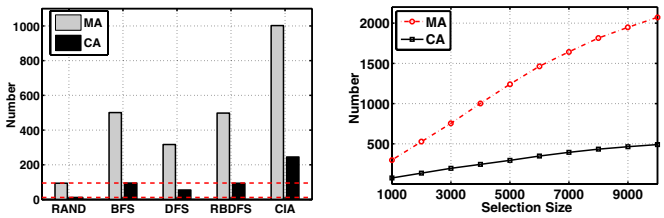
5.2.1 Evaluation on Dataset I

We first design five experiments to evaluate the effectiveness of our CIA based on Dataset I.

Different Selection Strategies. In this experiment, we start from the same seed set of N identified criminal accounts, which are randomly selected from 2,060 identified criminal accounts. Then, starting from this seed set, we use the following five strategies to select five different account sets with the same selection size of k from the dataset[5]: random search (RAND), breath-first search (BFS), depth-first search (DFS), random combination of breadth-first and depth-first search (RBDFS)[6], and CIA. From Figure 9(a), we can see that CIA can outperform all the other selection strategies. Specifically, CIA can infer 20.42 times as many CA and 10.66 *times* as many MA as that of using random selection strategy. Also, CIA can infer 2.58 times as many CA and around 2.00 times as many MA as that of using BFS, which can infer the second most CA. Also, CIA can perform much better than the naive algorithm that considering all accounts are possible criminal accounts. Specifically, CIA can correctly predict around 0.0625 criminal accounts and over 0.25 malicious affected accounts by selecting 1 account. However, the naive algorithm can only correctly predict 0.004 criminal accounts and 0.02 malicious affected accounts by selecting 1 account.

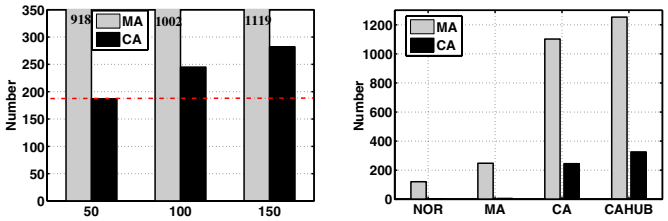
Different Selection Sizes: In this experiment, we start from 100 identified criminal account seeds and use CIA to infer criminal accounts by choosing different selection sizes of accounts, i.e., we evaluate our CIA by changing the values of k in the previous experiment. From Figure 9(b), we can see that when we select more accounts, we can infer more CA and MA, and the increase of CA and MA is sub-linear with the increase of the selection size.

Different Sizes of Seed Sets. In this experiment, we evaluate CIA by starting from different sizes of criminal seeds, i.e., we set different values of N. In this experiment, we also set $k = 4,000$. From Figure 10(a), we can see that when we increase the number of seeds, we can infer more criminal accounts while selecting the same size of accounts. This is because when we use more criminal seeds, we have more knowledge about the relationships among the criminal account community.



(a) Selection Strategies (b) Selection Sizes

Figure 9: Using different selection strategies and setting different selection sizes of accounts.



(a) Seed Size (b) Seed Type

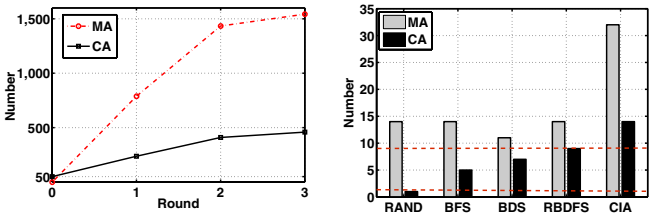
Figure 10: Striating from different sizes of seed sets and different types of seeds.

Different Types of Seeds. In this experiment, we evaluate CIA by using different types of accounts as the seeds. Specifically, we start from the same number (100) of randomly selected normal accounts (NOR) (posting no malicious tweets), malicious affected accounts (MA), criminal accounts (CA), and criminal hubs (CAHUB) and use CIA to select the same amount of 4,000 accounts. From Figure 10(b), we can find that starting from CAHUB and CA can predict much more CA and MA. Specifically, starting from CA, we can infer 245 CA and 1,102 MA, while starting from MA, we can infer 6 CA and 248 MA, and from NOR, we can infer 2 CA and 121 MA. This observation also validates that criminal accounts have stronger social relationships and semantic coordinations among themselves. Thus, it will be more effective to use known criminal accounts other than normal accounts as seeds to infer other criminal ones. In addition, we can also find that using CAHUB can even infer more CA and MA than using CA. That is also mainly because these criminal hubs have even more social relationships with other criminal accounts than criminal leaves.

Multiple Round Recursive Inference. In this experiment, we initially start from a small set of randomly selected 50 identified criminal accounts to recursively run CIA to infer criminal accounts. Specifically, during each round, we will combine previous round's seeds and identified criminal accounts correctly inferred in the previous round as new seeds to run CIA again. From Figure 11(a), we can find that even when we start from a small number of criminal accounts (50, which is around 2.4% of all CA in the dataset) within running 3 rounds of CIA, we can infer around 9 times more

[5]In this experiment, we choose $N = 100$ and $k = 4,000$.
[6]Specifically, when RBDFS traverses to an account, it will have a probability of 50% to make a breath-first or a depth-first search in the next step.

criminal accounts (500, which is around 22.3% of all CA). This observation shows that we can use CIA to recursively infer more criminal accounts by adding newly correctly inferred criminal accounts into the existing seed set.



(a) Recursive inference (b) Dataset II

Figure 11: Evaluation of multiple round recursive inference and Dataset II.

5.2.2 *Evaluation on Dataset II*

To decrease the effect of possible sampling bias in our analyzed dataset and to show the fact that the performance of CIA are reproducible, we also test CIA on another newly crawled dataset. Also, to guarantee the correctness of identifying criminal accounts, we first use Google Safe Browsing, a trustable blacklist, to collect malicious affected accounts. Then, we manually identify criminal accounts from those malicious affected accounts[7]. Then, we examine the effectiveness of CIA on newly crawled dataset by comparing different account selection strategies. Specifically, we start from only 10 identified criminal accounts and select 4,000 accounts by using each strategy. From Figure 11(b), we can also find that CIA can generate the best results. CIA can infer 13 more criminal accounts than that of using RAND.

Through all above experiments, we can find that CIA can be used to effectively infer unknown criminal accounts. Also, unlike most current work on detecting Twitter spammers based on machine learning techniques, which require extracting many features from all the accounts in the dataset, CIA mainly focuses on those accounts that have strong social relationships with existing known criminal accounts. In addition, CIA can be utilized to work as an early-stage monitoring and ranking algorithm to monitor those highly suspicious accounts, which may evolve to be criminal accounts later.

6. RELATED WORK

We discuss prior related work on Online Social Networks (OSN) by organizing them into two general categories.

6.1 Analysis of OSN Characteristics

Due to the great popularity of the OSNs, many studies have analyzed OSN characteristics. Mislove *et al.* present a large-scale measurement study on the structure of multiple OSNs including Flickr, YouTube, LiveJournal, and Orkut [28]. Kwak *et al.* have shown a comprehensive and quantitative study on Twitter accounts' behavior [25]. Yardi *et al.* analyze social behavior and network structural differences between spam Twitter accounts and legitimate ones by analyzing a specific spam campaign [36]. However, due to the limitation of the analyzed dataset, it fails to reveal in-depth structural differences to answer the questions mentioned in that work, e.g., "Do spammers follow one another to boost their follower count?". Wang *et al.* use Twitter to study the unbiased sampling algorithm for directed social graphs [34]. Cha *et al.* utilize different metrics to measure the user influence on Twitter [18]. Galuba *et al.* focus on characterizing and modeling the information cascades formed by individual URL mentions in the Twitter follower graph [20]. Castillo *et al.* design automatic methods for assessing the credibility of a given set of tweets [17]. Metaxas *et al.* analyze political community behavior and the spread of political opinions on Twitter [27], and Ratkiewicz *et al.* analyze the spread of Astroturf memes on Twitter [30].

6.2 Mitigation of OSN Criminal Accounts

In addition, since spam and attacks are so rampant in the OSNs, many researchers have studied detecting OSN criminal accounts. A framework to detect tag spam in tagging systems is proposed in [24]. This work prevents the attackers who desire to increase the visibility of an object from fooling the search mechanism. Benevenuto *et al.* [15, 16] utilize machine learning techniques to identify video spammers on YouTube. Gao *et al.* present a study on detecting and characterizing social spam campaigns in Facebook [21]. Meanwhile, most Twitter criminal account detection work can be classified into two categories. The first category of work, such as [26, 14, 33, 32], utilizes machine learning techniques to classify legitimate accounts and criminal accounts according to their collected training data and their selections of classification features. The second category of work (e.g., [22]) detects and analyzes malicious accounts by examining whether URLs or domains posted in the tweets are labeled as malicious by public URL blacklists or domain blacklists.

Compared with previous work, our work focuses more on the analysis of cyber criminal ecosystem – investigating inner social relationships in the criminal account community and outer relationships between criminal accounts and criminal supporters – to deeply understand how criminal accounts survive and mix into the whole Twitter space. Thus, our study is a valuable supplement to the previous work.

7. LIMITATIONS AND FUTURE WORK

We acknowledge that our analyzed dataset may contain some bias. Also, the number of our analyzed criminal accounts is most likely only a lower bound of the actual number in the dataset, because we only target on one specific type of criminal accounts due to their severity and prevalence on Twitter. However, it is extremely challenging to obtain an ideal, unbiased dataset with perfect ground truth. In addition, to reduce possible data sampling bias, we crawled two datasets at very different time to evaluate the performance of our CIA. We also believe that even though the exact values of some metrics used in our work may vary a little bit when using different sample datasets, our major conclusions and insights will likely still hold.

We also acknowledge that our validations on some possible explanations proposed in this work may be not absolutely rigorous, due to the difficulties in thoroughly obtaining criminal accounts' social actions or motivations. However, we believe that our first-in-its-kind analysis of those phenomenon still provides great values and opens a new door to better understand the cyber criminal ecosystem on Twitter.

In our future work, we will design and test more crawling strategies and crawl more data. We also plan to further

[7]We acknowledge that the numbers of CA and MA are the low bound of real numbers in the dataset, because we can not detect all CA and MA by simply using GSB itself.

deeply analyze the differences between criminal accounts' relationship graph and that of normal accounts. In addition, we plan to design a full detection system by combining our CIA algorithm and other detection features.

8. CONCLUSION

In this paper, we present an empirical analysis of the cyber criminal ecosystem on Twitter. We provide in-depth investigation on inner and outer social relationships. We observe two findings in the cyber criminal community and reveal the characteristics of three representative categories of criminal supporters. Spurred by defense insights originating from these analyses, we design an effective algorithm to infer more criminal accounts by starting from a seed set of known criminal accounts and exploiting the properties of their social relationships and semantic correlations.

9. REFERENCES

[1] Automatically Follow Back. http://autofollowback.com.

[2] Calculate Twitter Reputation. http://www.thekirankumar.com/blog/tag/calculate-twitter-reputation/.

[3] Capture HPC. https://projects.honeynet.org/capture-hpc.

[4] Google Safe Browsing API. http://code.google.com/apis/safebrowsing/.

[5] KOOBFACE: Inside a Crimeware Network. http://www.infowar-monitor.net/reports/iwm-koobface.pdf.

[6] Lady Gaga Falls Prey to Rogue Twitter Attack. http://mashable.com/2011/04/28/lady-gaga-twitter-attack/.

[7] Purchase Twitter Friends. http://www.purchasetwitterfriends.com/.

[8] The Twitter Rules. http://help.twitter.com/entries/18311-the-twitter-rules.

[9] Twitter accounts spreading malicious code. http://www.net-security.org/malware_news.php?id=1554.

[10] Twitter-based Botnet Command Channel. http://asert.arbornetworks.com/2009/08/twitter-based-botnet-command-channel/.

[11] Twitter Streaming API. https://dev.twitter.com/docs/streaming-api.

[12] Twitter vulnerability allows cyber criminals to spread spam. http://www.one.com/en/web-hosting-news/website/twitter-vulnerability-allows-/cyber-criminals-to-spread-spam-links\$800076628.htm.

[13] Twitter's Following Limits. http://support.twitter.com/groups/32-something-s-not-working/topics/117-following-problems/articles/66885-i-can-t-follow-people-follow-limits.

[14] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting Spammers on Twitter. In *Collaboration, Electronic messaging, Anti-Abuse and Spam Conffference (CEAS)*, 2010.

[15] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, C. Zhang, and K. Ross. Identifying Video Spammers in Online Social Networks. In *Int'l Workshop on Adversarial Information Retrieval on the Web (AirWeb'08)*, 2008.

[16] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, C. Zhang, and K. Ross. Detecting Spammers and Content Promoters in Online Video Social Networks. In *ACM SIGIR Conference (SIGIR'09)*, 2009.

[17] C. Castillo, M. Mendoza, and B. Poblete. Information Credibility on Twitter. In *International World Wide Web Conference, (WWW'11)*, 2011.

[18] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Int'l AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.

[19] G. Danezis and P. Mittal. SybilInfer: Detecting Sybil Nodes using Social Networks. In *16th Annual Network and Distributed System Security Symposium*, 2009.

[20] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer. Outtweeting the Twitterers - Predicting Information Cascades in Microblogs. In *Usenix Workshop on Online Social Networks, (WOSN'10)*, 2010.

[21] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao. Detecting and Characterizing Social Spam Campaigns. In *Proceeding of ACM SIGCOMM IMC(ICM'10)*, 2010.

[22] C. Grier, K. Thomas, V. Paxsony, and M. Zhangy. @spam: The Underground on 140 Characters or Less. In *ACM Conference on Computer and Communications Security (CCS)*, 2010.

[23] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *Journal of the ACM, Vol.46, No. 5, pp. 604-632*, 1999.

[24] G. Koutrika, F. Effendi, Z. Gyongyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *Int'l Workshop on Adversarial Information Retrieval on the Web (AIRWeb'07)*, 2007.

[25] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Int'l World Wide Web (WWW '10)*, 2010.

[26] K. Lee, J. Caverlee, and S. Webb. Uncovering Social Spammers: Social Honeypots + Machine Learning. In *ACM SIGIR Conference (SIGIR)*, 2010.

[27] P. Metaxas and E. Mustafaraj. Prominence in minutes: Political speech and real-time search. In *In Proceedings of the Web Science 2010*, 2010.

[28] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Outtweeting the Twitterers - Predicting Information Cascades in Microblogs. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, (IMC'07)*, 2007.

[29] D. Pelleg and A. Moore. X-Means: Extending K-means with Efficient Estimation. In *International Conference on Machine Learning*, 2000.

[30] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, S. Patil, A. Flammini, and F. Menczer. Detecting and Tracking the Spread of Astroturf Memes in Microblog Streams. In *Proceedings of 5th International Conference on Weblogs and Social Media*, 2011.

[31] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. In *Information Processing & Management*, 1998.

[32] G. Stringhini, S. Barbara, C. Kruegel, and G. Vigna. Detecting Spammers On Social Networks. In *Annual Computer Security Applications Conference*, 2010.

[33] A. Wang. Don't follow me: spam detecting in Twitter. In *Int'l Conferene on Security and Cryptography (SECRYPT)*, 2010.

[34] T. Wang, Y. Chen, Z. Zhang, P. Sun, B. Deng, and X. Li. Unbiased Sampling in Directed Social Graph. In *ACM Special Interest Group on Data Communication, (SIGCOMM'10)*, 2010.

[35] C. Yang, R. Harkreader, and G. Gu. Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. In *In Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection (RAID'11)*, 2011.

[36] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd. Detecting spam in a Twitter network. In *First Monday, 15(1)*, 2010.

[37] H. Yu, M. Kaminsky, P. Gibbons, and A. Flaxman. SybilGuard: Defending Against Sybil Attacks via Social Networks. In *Proceedings of ACM SIGCOMM Conference*, 2006.

[38] J. Zhang, P. Porras, and J. Ullrich. Highly predictive blacklisting. In *17th USENIX Security Symposium (USENIX Security'08)*, 2008.