Analyzing Spammers’ Social Networks for Fun and Profit

A Case Study of Cyber Criminal Ecosystem on Twitter

Presentation: Michelle Hanne S. Andrade
SUMMARY

I. INTRODUCTION
II. RESEARCH GOAL AND DATASET
III. INNER SOCIAL RELATIONSHIPS
IV. OUTER SOCIAL RELATIONSHIPS
V. INFERRING CRIMINAL ACCOUNTS (CIA)
VI. RELATED WORK
VII. LIMITATIONS AND FUTURE WORK
VIII. CONCLUSION
I- INTRODUCTION

- Cyber criminals have utilized Twitter as a new platform to conduct their malicious behavior including sending spam and phishing scams, spreading malware, hosting botnet command and control (C&C) channels, and launching other underground illicit activities.

- Criminal accounts’ social relationships can aid them in increasing the visibility of their malicious content – thus in obtaining more victims.
I- INTRODUCTION

- Can evade existing detection approaches such as “Twitter Rules” and break through Twitter’s “Follow Limit Policy”, while maintaining their high visibility.

**Twitter Rules:**
“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.”
I- INTRODUCTION - Questions

- 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?
I- INTRODUCTION

cyber criminal ecosystem on Twitter, containing criminal account community composed of criminal accounts, and criminal supporter community composed of those accounts outside the criminal account community who have close friendships (following relationships) with criminal accounts, defined in our work as criminal supporters.

Figure 1: Structure of the cyber criminal ecosystem.
II- RESEARCH GOAL AND DATASET

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, who mainly post malicious URLs linking to malicious content with an intention to compromise users’ computers or privacy.
II- RESEARCH GOAL AND DATASET

Was analyzed the dataset from previous Twitter spam account detection study which was crawled by tapping into Twitter’s Streaming API, from April 2010 to July 2010.

<table>
<thead>
<tr>
<th>Twitter accounts</th>
<th>485,721</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets</td>
<td>14,401,157</td>
</tr>
<tr>
<td>URLs</td>
<td>5,805,351</td>
</tr>
</tbody>
</table>
II- RESEARCH GOAL AND DATASET

- To analyze criminal accounts, we also use the results from that previous study, 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) and a high-interaction client honeypot, implemented using Capture-HPC.
III- INNER SOCIAL RELATIONSHIPS

**Visualizing Relationship Graph**

The criminal relationship is a directed graph $G = (V,E)$

- each criminal account as a node
- each follow relationship as a directed edge

In dataset, the criminal relationship graph consists of 2,060 nodes and 9,868 directed edges
III-INNER SOCIAL RELATIONSHIPS

The giant connected component contains 954 nodes.

By further breaking down the graph, we can obtain **8 weakly connected components containing at least three nodes and 521 isolated nodes**.
III- INNER SOCIAL RELATIONSHIPS

Revealing Relationship Characteristics

1) Criminal accounts tend to be socially connected, forming a small-world network.

Three metrics graphs were used to validate that criminal accounts tend to socially connect with each other: 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:

\[
\frac{|E|}{|V| (|V| - 1)}
\]

This metric measures how closely a graph is to be a complete graph.
After calculating the graph density for both our sample criminal relationship and a public entire Twitter snapshot containing 41.7 million users and 1.47 billion edges, we find that the graph density of our sample criminal relationship graph, which is $2.33 \times 10^{-3}$, is much higher than that of the Twitter snapshot, which is $8.45 \times 10^{-7}$. This shows that the criminals have closer relationship than regular Twitter users.
Reciprocity is represented by the number of bi-directional links to the number of outlinks.

55% normal accounts in our crawled graph (containing around 500K nodes)

Around 20% of criminal accounts’ values of reciprocity in the criminal graph are nearly 1.0

(a) Reciprocity

95% criminals accounts have the reciprocity higher than 0.2 in the criminal graph
III- INNER SOCIAL RELATIONSHIPS

- **Average Shortest Path Length** is defined as the average number of steps along the shortest paths for all possible pairs of graph nodes.
- Compared with the average path length of a sample data set with 3,000:

<table>
<thead>
<tr>
<th>Account Type</th>
<th>Average Path Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legitimate Twitter accounts</td>
<td>4.12</td>
</tr>
<tr>
<td>Criminal relationship</td>
<td>2.60</td>
</tr>
</tbody>
</table>

From the above analysis, we can find that criminal accounts have strong social connections with each other.
III- INNER SOCIAL RELATIONSHIPS - Possibles

- 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.
III- INNER SOCIAL RELATIONSHIPS - Possibles

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.
III- INNER SOCIAL RELATIONSHIPS

2- Compared with criminal leaves, criminal hubs are more inclined to follow 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.

Criminal Following Ratio (CFR), which is the ratio of the number of an account’s criminal-followings to its total following number.
III- INNER SOCIAL RELATIONSHIPS

- Around 80% of criminal hubs’ CFRs are higher than 0.1.
- Only 20% of criminal leaves are higher than 0.1.
- 60% of criminal leaves’ CFRs are lower than 0.05.
III- INNER SOCIAL RELATIONSHIPS

- Was designed 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.
III- INNER SOCIAL RELATIONSHIPS

Around 80% of criminal hubs’ SFRs are higher than 0.4.

5% of criminal leaves are higher than 0.4.

This observation reflects that compared with criminal leaves, criminal hubs’ followers share more followers with their criminal-followings.
III- INNER SOCIAL RELATIONSHIPS

Compared to the Bee Community

Criminal leaves, like bee workers, mainly focus on collecting pollen (randomly following other accounts to expect them to follow back.

Criminal hubs, like bee queens, mainly focus on supporting bee workers and acquiring pollen from them (following leaves and acquiring their followers’ information).
IV- OUTER SOCIAL RELATIONSHIPS

Criminal Supporters: They are 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.
Extracting Criminal Supporters

They design a *Malicious Relevance Score Propagation Algorithm (Mr.SPA)* to extract criminal supports.

Mr.SPA assigns a malicious relevance score (MR score) to each Twitter account, measuring how closely this account follows criminal accounts.
IV- OUTER SOCIAL RELATIONSHIPS

The MR score was measured 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.
IV- OUTER SOCIAL RELATIONSHIPS

To formalize the above intuitions, they built a malicious relevance graph $G = (V,E)$ to model the support relationship.

In this graph, they considered 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 $j$. 
IV- OUTER SOCIAL RELATIONSHIPS

MR Score Initialization: initial score \( M^0_i \) 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 \( m_i \). For other accounts, the score is initialized to zero.

MR Score Propagation: the three below heuristic.
IV- OUTER SOCIAL RELATIONSHIPS

- **Policy 1: MR Score Aggregation.** An account’s score should sum up all the scores inherited from the accounts it follows.

\[
M(C_1) = M_1 \\
M(C_2) = M_2 \\
M(A) = M_1 + M_2
\]

A follows both criminal accounts C1 and C2, the score of A is the sum of the malicious scores of C1 and C2.
IV- OUTER SOCIAL RELATIONSHIPS

- **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 $\alpha$ according to their social distances, where $0 < \alpha < 1$.

  When $A_1$ is one hop away from a criminal account $C$, we assign it a dampening factor of $\alpha$, where $0 < \alpha < 1$. When $A_2$ is two-hop away, $A_2$ will get a dampening factor of $\alpha \cdot \alpha = \alpha^2$. 

\[
M(C) = M \\
M(A_1) = \alpha \times M \\
M(A_2) = \alpha \times \alpha \times M
\]
IV- OUTER SOCIAL RELATIONSHIPS

- **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.

  A₁ and A₂ 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 $(i, j) \in E$. 
IV- OUTER SOCIAL RELATIONSHIPS

- 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

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

- Notation: $l_{ij} = \{0, 1\}$ to indicate whether $(i, j) \in E$ (if $(i, j) \in E$, $l_{ij} = 1$; otherwise, $l_{ij} = 0$).
IV- OUTER SOCIAL RELATIONSHIPS

- In addition, with the consideration of each node’s historical score record, at each step \( t(t > 0) \), we add an initial score bias \((1 - \alpha)\). In our experiment, we set \( \alpha = 0.85 \), since it is widely used in the random-walk.

\[
\overrightarrow{M_t^t} = \alpha \cdot \overrightarrow{I} \cdot \overrightarrow{M^{t-1}} + (1 - \alpha) \cdot \overrightarrow{M^0} \quad (t > 0)
\]

- Malicious score column vector for all nodes at the step \( t \).
- Initial MR Score vector for all nodes.
When the score vector converges after several propagation steps, we can obtain final MR scores for all nodes. To find an acceptable threshold, we first use $k$-means algorithm to cluster accounts based on their MR scores.

They observed that most accounts have relatively small scores and are grouped into one single cluster.

With this observation, they chose the highest score of the account in that cluster as the threshold. Then, we output 5,924 criminal supporters, whose MR scores are higher than the threshold,
IV- OUTER SOCIAL RELATIONSHIPS

Characterizing Criminal Supporters

After extracting criminal supporters we observe three representative categories of supporters (social butterflies, social promoters, and dummies) according to our defined thresholds.

- **Social Butterflies** are those accounts that have extraordinarily large numbers of followers and followings. These accounts build a lot of social relationships with other accounts without discriminating those accounts’ qualities.
IV- OUTER SOCIAL RELATIONSHIPS

- 2,000 following as a threshold in terms of Twitter’s Follow Limit Policy. They found 3,818 social butterflies.

- The 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.

- To validate this hypothesis they did an experiment with 10 twitter accounts. The fast speed with which social butterflies went back represents 47.8%.
IV- OUTER SOCIAL RELATIONSHIPS

- **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.

- Extract those social promoters whose URL ratios 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. **508 social promoters.**
IV- OUTER SOCIAL RELATIONSHIPS

- **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.

- Extract those social promoters whose URL ratios 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. **508 social promoters.**
IV- OUTER SOCIAL RELATIONSHIPS

- Promoters may become criminal supporters by unintentionally following criminal accounts, because following other accounts without considerations of those accounts’ quality.

- They used a heuristic method to validate our hypothesis. From the repetition of domain names in the URL of the Promoters was calculated Entropy.

\[- \sum_{i=1}^{N} p_i \ln p_i\]

- denotes the number of distinct domain names
- denotes the ratio of the occurrences of the i-th distinct domain name to the total number of domain names.
IV- OUTER SOCIAL RELATIONSHIPS

Around 80% social promoters have the domain name entropy lower than 1.0.

Around 45% of all accounts have the domain name entropy lower than 1.0.

Around 40% social promoters’ domain name entropy are zero, which implies that all their URLs have the same domain names.
IV- OUTER SOCIAL RELATIONSHIPS

- **Dummies** are those Twitter accounts who post few tweets but have many followers.
- The 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.
- Analyzed 81 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.
This dummy account steals victims’ email addresses through claiming to help people earn money. However, the dummy account sends email spam.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- 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.
- They proposed a **Criminal accounts Inference Algorithm (CIA)** to infer more criminal accounts by exploiting criminal accounts’ social relationships and semantic coordinations.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Design of CIA**: 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.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- The CIA algorithm integrates with Mr.SPA to quantify the closeness of social relationships. Integrated a metric, Semantic Similarity score (SS score).

- To calculate SS score, 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.

Higher SS score between two accounts implies that they have stronger semantic coordinations.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- They built a malicious relevance graph, denoted as
  \[ G = (V,E) \]
  each follow relationship denotes a directed edge
each account denotes a vertex

Assigned 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. Then, the weight \( WS(i, j) \) of the edge \( e_{ij} \) can be calculated as

\[
WS(i, j) = \frac{SS_{ij}}{\sum_{e_{kj} \in E} SS_{kj}}
\]

When the score vector converges after several propagation steps, we infer those accounts with high malicious scores as criminal accounts.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Evaluation of CIA:** The CIA was evaluated based on two different datasets - **Dataset I** refers to the one with around half million accounts from our previous study. **Dataset II** contains another new crawled 30K accounts by starting from 10 newly identified criminal accounts and using breath-first search (BFS) strategy.

- They adopted similar metrics to “Hit Count” measure CIA’s effectiveness rather than using false positive and false negative rate.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Evaluation on Dataset I:** In this experiment, they started from the same seed set of $N$ identified criminal accounts, which are randomly selected from 2,060 identified criminal accounts. They used the following five strategies to select five different accounts: random search (RAND), breath-first search (BFS), depth-first search (DFS), random combination of breadth-first and depth-first search (RBDFS).

In this experiment, we choose $N = 100$ and $k = 4,000$

- $N$: seed set
- $K$: size of dataset
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- CIA can outperform all the other selection strategies.

\[ \text{CIA can infer } 20.42 \text{ times as many CA and } 10.66 \text{ times as many MA as that of using random selection strategy} \]

\[ \text{CA} = \text{the number of correctly inferred criminal accounts.} \]
\[ \text{MA} = \text{the number of malicious affected accounts.} \]
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- 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. However, the **naive algorithm** can only correctly predict **0.004** criminal accounts and **0.02** malicious affected accounts.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Different Selection Sizes:** they evaluated our CIA by changing the values of $k$ in the previous experiment.

The increase of CA and MA is sub-linear with the increase of the selection size.

(b) Selection Sizes
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Different Sizes of Seed Sets**: CIA by starting from different sizes of criminal seeds, i.e., we set different values of N. 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) Seed Size
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Different Types of Seeds:** started from the same number (100) of randomly selected normal accounts (NOR), malicious affected accounts (MA), criminal accounts (CA), and criminal hubs (CAHUB) and use CIA to select the same amount of 4,000 accounts.

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.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

Multiple Round Recursive Inference: start from a small set of randomly selected 50 identified criminal accounts to recursively run CIA to infer criminal accounts. During each round, are combined previous round’s seeds and identified criminal accounts correctly inferred in the previous round as new seeds to run CIA again.

![Graph showing recursive inference]

- Started from a small number of criminal accounts (50, which is around 2.4% of all CA in the dataset)
- running 3 rounds of CIA, we can infer around 9 times more
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- **Evaluation on Dataset II**: They examined the effectiveness of CIA on newly crawled dataset by comparing different account selection strategies. Started from only 10 identified criminal accounts and select 4,000 accounts by using each strategy.

![Graph showing account selection strategies](image)

- CIA can infer 13 more criminal accounts than that of using RAND.

- CIA can generate the best results.
V- INFERRING CRIMINAL ACCOUNTS (CIA)

- 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.
VI- RELATED WORK

Most Twitter criminal account detection work can be classified into two categories.

- The first category of work, 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 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.
VI- RELATED WORK

Compared with previous work, this 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.
VII- LIMITATIONS AND FUTURE WORK

- The dataset may contain some bias.
- 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.
- 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.
VII- LIMITATIONS AND FUTURE WORK

- Future work, intend design and test more crawling strategies and crawl more data, further deeply analyze the differences between criminal accounts’ relationship graph and that of normal accounts.

- Plan to design a full detection system by combining our CIA algorithm and other detection features.
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.