What They Do in Shadows: Twitter Underground Follower Market

Anupama Aggarwal, Ponnurangam Kumaraguru
Indraprastha Institute of Information Technology, Delhi (IIIT-D)
Cybersecurity Education and Research Centre (CERC), IIIT-Delhi
Email: {anupamaa,pk}@iiitd.ac.in

***Abstract*—Internet users and businesses are increasingly using online social networks (OSN) to drive audience traffic and increase their popularity. In order to boost social presence, OSN users need to increase the visibility and reach of their online profile, like - Facebook likes, Twitter followers, Instagram comments and Yelp reviews. For example, an increase in Twitter followers not only improves the audience reach of the user but also boosts the perceived social reputation and popularity. This has led to a scope for an underground market that provides followers, likes, comments, etc. via a network of fraudulent and compromised accounts and various collusion techniques.**

In this paper, we landscape the underground markets that provide Twitter followers by studying their basic building blocks - merchants, customers and phony followers. We charecterize the services provided by merchants to understand their operational structure and market hierarchy. Twitter underground markets can operationalize using a premium monetary scheme or other incentivized *freemium* schemes. We find out that freemium market has an oligopoly structure with few merchants being the market leaders. We also show that merchant popularity does not have any correlation with the quality of service provided by the merchant to its customers. Our findings also shed light on the characteristics and quality of market customers and the phony followers provided by underground market. We draw comparison between legitimate users and phony followers, and find out key identifiers to separate such users. With the help of these differentiating features, we build a supervised learning model to predict suspicious following behaviour with an accuracy of 89.2%.

I. Introduction

Social media presence has become vital for businesses for lead generation, and users to increase their popularity amongst their friends network. In order to enhance and maintain social media presence, users need to generate a following for their social profile, such as - likes on the Facebook page, followers on Twitter and comments on Instagram post. Recent studies have indicated the growth of underground markets for the purchase of Twitter followers, Facebook likes, Instagram followers and Yelp reviews [1], [3], [12], [13], [16], [17]. Users subscribe to services of underground markets to artificially boost their social media presence and influence. An increase in Twitter followers not only improves the audience reach of the user but also boosts her perceived social reputation and popularity. Rising demand of Twitter followers has led to the growth of an underground industry that caters to users' need for quick followers. We refer to this underground industry as *follower market* and to their operators as *follower merchants*.

Underground market has constantly evolving techniques to provide phony followers like (i) selling fraudulent accounts, i.e., pseudonym accounts which act as fake followers; (ii) using compromised accounts where the malware on user's machine or compromised credentials cause her to follow customer's account without user's knowledge; (iii) leveraging collusion networks where customers are incentivized to become part of the follower network [13], [16]. To understand the structure and characteristics of Twitter *follower market*, we focus on its basic building blocks viz. (i) *follower merchants*, (ii) *customers*, i.e. the user who take services from these merchants, and (iii) *phony followers*, i.e., the followers provided as a service to a customer by the merchant. An artificially inflated follower count can give the user a veneer of importance and popularity in Twittersphere.

This study landscapes the Twitter follower market. Characterization and analysis of 60 *freemium* and 57 *premium* markets shed light on (i) structure of the follower merchants, (ii) quality of customers and (iii) key identifiers to distinguish between phony follower accounts and legitimate users. In particular, we present following contributions:

First, we conducted a longitudinal study to characterize Twitter *follower merchants*. In order to identify the most popular merchants and market leaders, we introduced the idea of *Quality of Service* (QoS) for the Twitter follower markets. QoS is an important parameter to judge the overall performance of a service, in our case, the follower merchants [5]. As discussed in Section III-B, we defined a metric for QoS, which takes into account the services promised by the merchant, expectation by the customer and difference between the two. Using the QoS and popularity metric, we were able to highlight a hierarchy of follower merchants in the underground market and show that this market exhibits an *oligopoly* structure.

Second, we assess the *customers* taking services from the follower market. We characterized customers of various merchants on the basis of their social reputation and profile attributes. We observed that customers lying on the higher strata of quality take services of freemium merchants.

Lastly, We present an anatomy of the purchased Twitter followers. We characterized profile attributes and behavioural features of purchased followers. We identifed key indicators to distinguish between suspicious following behaviour from that of legitimate Twitter users. We used these identifiers and built a supervised learning mechanism which detects suspicious following behaviour with an accuracy of 89.2%.

II. Related Work

Researchers have shown that miscreants use several strategies to monetize spam and other malicious activities [9]. There exists a large underground market which sells specialized services and products like fraudulent accounts [14], [16], solving CAPTCHA [10], pay-per-install [6], and writing fake reviews or website content [11], [18]. Social media users take such services to increase their online presence. For example, on Twitter, users attempt to gain followers in order to boost their popularity [7]. Underground markets are a threat to the QoS and are generating a revenue of about $360 million per year from sale of fake Twitter followers [2]. In this paper, we present a comprehensive study of Twitter follower market to understand how it operates and assess the QoS [5] and percieved gain by the use of phony follower merchants.

Recent studies have shown that merchants often create fake accounts to deliver services like phone verified email accounts [15], Twitter followers [16] and Facebook Likes [4]. Researchers show that such fraudulent accounts can be detected at the time of account creation by merchants by finding patterns in account naming convention and registration process [16]. In this paper, we focus on phony followers of the customers. However, we do not limit our study to only fraudulently created accounts but also study compromised accounts exhibiting phony follow behaviour. Researchers have modelled suspicious Twitter following behaviour by identifying difference in follow pattern from the majority [8]. Previous studies highlight the unfollow dynamics of the victim customer accounts whose credentials are compromised by merchants [13]. In this paper, we study not only compromised users but also the legitimate users part of the collusion follower network.

III. Background

A. Freemium and Premium Markets

Premium Market Under the *premium* scheme, customers have to pay money to the merchants in order to gain followers. Customer provides a Twitter username to the merchants for which she wants to increase the follower count, and purchases a specific package (e.g. package - 1000 followers for $3). Premium market allows a customer to purchase bulk followers for not just himself but any Twitter user. The merchants under the premium scheme, either simply sell followers in exchange of money, or also require the customer to provide her Twitter account's password so that the customer can be made part of the collusion phony follower network. Note that in both the cases, merchants require monetary payment from the customer.

Freemium Market *Freemium* market scheme lets the customer gain followers without any monetary payment. However, in return, the customer needs to authorize merchant's Twitter application that enables the merchant to include customer in the collusion phony follower network. Once the customer authorize merchant's app, she starts gaining followers within minutes. The merchant app includes various permissions like - *see who you follow, follow new people, update your profile*, and *post tweets for you*. These permissions enable the merchant to make the customer follow other Twitter users (which may be other customers of the merchant) and also post promotional tweets on her behalf.

Figure 1 summarizes how the two market schemes operate. It can be seen that freemium market operate primarily by leveraging collusion networks. This causes even legitimate users to exhibit phony follow behavior in return of bulk followers. Customers who provide their passwords to merchants under the premium scheme are at a high risk of being compromised. Once compromised, these accounts can be fully controlled by the merchants and used for following other customers, spamming or sending promotional tweets as and when required.

B. Quality of Service

Twitter follower merchants lay down various terms of services to the customers who subscribe to them. As of now, we do not understand to what extent these merchants violate their terms of services, who are the market leaders and which merchants have maximum penetration in the market. To answer all these questions, we use existing literure on consumer research by Bolten et al. which provides a conceptual framework to model customer's assessment of service quality and value [5]. We apply the formulization of Quality of Service (QoS) to the underground follower merchants to understand the quality and hierarchy of merchants. Researchers have shown that the measure of QoS can be based on performance of the service, expectation by the customer and the gap between these two parameters. Based on this, QoS can be defined as a function q –

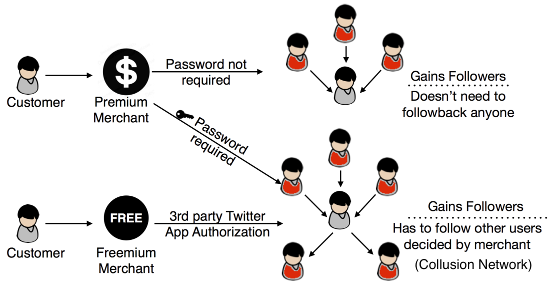


Fig. 1: Different market schemes of follower market. Merchants can enforce monetary payment or use Twitter application to incentivize the customer. In both schemes, merchant can leverage collusion networks to include customers into the phony follower network.

$$\textbf{QoS} = q(PERFORM, EXPECT, DISCONFIRM) \quad (1)$$

where PERFORM is a vector of the performance of a merchant based on several terms of services $\{SA_1...SA_k\}$, i.e.,

$$PERFORM = p_k(SA_k) \quad (2)$$

EXPECT is a vector which describes the prior expectations of the customer for each term of service $\in \{SA_1...SA_k\}$. DISCONFIRM is a vector describing the amount of discrepancy between performance of the merchant and expectation of the customer for each of the terms of services.

IV. Data Collection Methodology

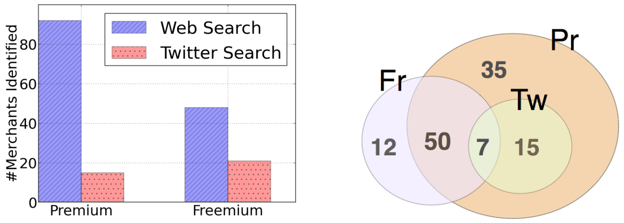
We collected data from both *premium* as well as *freemium* merchants. Since we want to landscape the merchants, customers and followers, we underwent a three step process to collect data.

a) ***Merchant Identification****:* One of the building blocks of underground follower market is the *merchants*, i.e., the operators who provide phony followers to customers. In order to identify the merchant websites, we used search engine queries and filtered results of Twitter search. Table I shows the list of keywords which we used to identify merchant websites from premium and freemium market. We limited our search space to web engine (Google and Bing) results, tweets and Twitter user profile descriptions which match the keywords or a combination of keywords in the category.

TABLE I: shows keywords which are used to identify merchants. Search engine results are directly used after manual filtering. Twitter search results reveal promotional tweets as well as user profiles which reveal more merchant websites.

	Keywords	
	Premium	**Freemium**
Web Search	buy, Twitter, followers, bulk, increase, order, grow, gain, more, real, cheap	latest, riders, free, followers, Twitter, 'get more'
Twitter Search	recommend to gain, gain followers, gamer follow train, wana gain followers, need more followers, cheap followers	

We manually cleaned the web search results to identify merchant websites and group them into two categories - 'freemium' and 'premium'. Some merchants offer both *freemium* and *premium* schemes, therefore, they are grouped into both the categories. Figure 2(a) shows the distribution of websites we obtain by Twitter search and web engine queries. Freemium merchants have a larger tweet presence than premium merchants. We posit that this is because freemium merchants primarily operate using collusion network and sending promotional tweets. Twitter search also reveals that few merchants maintain a Twitter profile to gain audience. Figure 2(b) shows the distribution of merchants in both categories and indicates the merchants who maintain a Twitter profile in our dataset. We identifed 69 *freemium* merchants and 107 *premium* merchants. Out of the 69 freemium merchants we identifed, only 60 merchant websites publically displayed the latest customers on their websites. Also, we purchased followers from 57 of the premium merchants in our dataset.



(a) Merchants located by Twitter search and web engine search in each category.

(b) Pr=premium, Fr=freemium and Tw=merchants which have a Twitter profile.

Fig. 2: Distribution of premium and freemium merchants in our dataset. We use Twitter search and web search queries to locate a merchant. We find that some merchants have a Twitter profile and offer both premium and freemium services.

*b) **Phony Follower Data Collection**:* To identify large number of phony followers, we subscribed to services of premium and freemium merchants. For this task, we created dummy Twitter accounts to get followers from freemium and premium merchants. Table II shows services provided by the merchants from where we received the phony followers. In case of freemium merchants, we used dummy accounts to subscribe to each of the merchants; we used one dummy account per merchant. We authorized the Twitter application of each merchant website, in return of which we recieved followers. The followers were added to our dummy accounts at an average rate of 84 unique followers per hour as shown in Table III and overall we received 82,808 followers. For premium merchants, we used one dummy account per merchant to gain followers. We purchased the basic package from each of the merchants and obtained 87,458 followers.

TABLE II: Shows the types of services provided by the merchants from where we recieved our dataset of phony followers.

	Model	#Merchants	#Followers
Freemium	Only Freemium Freemium + Premium	12 57	82,808
Premium	Only Premium Freemium + Premium	5 52	87,458

As shown in Table III, we further observe that there is a lot of variation in number of followers obtained in case of freemium market, though we subscribed to similar services. Likewise, in case of premium market, we ordered minimum 1,000 followers but received as low as 738 followers. This indicates clear difference in operations of various merchants, which we further discuss later.

*c) **Market Customer Ground Truth Data**:* Customer users of premium merchants are not disclosed by the merchant websites. However, in case of freemium merchants, the latest customers are often displayed on the website with a link to their Twitter profile, and the list is refreshed after every few minutes. In order to collect the ground truth data of customers, we collected data from 60 such freemium merchant websites which provide link to their latest customers and obtained 171k unique customer profiles by taking hourly snapshots. Table IV describes the customer dataset in more detail.

TABLE III: Shows the number of phony followers collected from each kind of market. The last column shows the distribution of each parameter for all the merchants of specified category.

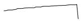
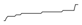
		Mean	Median	Min	Max	Distribution
Fr	Followers/Hr	84	85	60	105	
	Followers	1,505	1,524	678	2,030	
Pr	Cost/1000 Followers	$8.4	$8	$3	$14	
	Followers	1,590	1,607	738	2,095	

TABLE IV: Dataset description of customers located in freemium market over a period of 4 months by scraping 60 merchant websites.

Data Collection Timeframe	2014-07-18 - 2014-11-18
Number of Snapshots	2,870
Number of Merchants	60
Number of Customers Located	171,234
Verified Customers	10

Customers of premium merchants are not disclosed. Hence, we limit our study to the customers of freemium market and study their behaviour.

V. Landscaping Twitter Follower Market

A. Twitter Follower Merchants

Merchants are the market operators which provide phony followers to their customers. We recall that merchants can offer *premium*, *freemium* or both the schemes to their customers.

1) Merchants violate their promises: The merchants of underground Twitter follower market offer various guaranteed services to customer at the time of subscription. Many merchants promise services like authentic followers, moneyback guarantees, quick followers and follower retention which encourage the customers to either purchase bulk followers from premium merchants or subscribe to freemium merchants. Table V shows the list of most common promises made by the merchants. Though these promises and guarantees seem lucrative, and hence attract a lot of customers, merchants often violate these services.

Lack of follower retention: Most of the preemium merchants provide follower retention policy, i.e., they state that - *"if you loose any number of followers...we'll refill the page with the lagging followers, at absolutely free of cost"*. Therefore, we expect that customers will always have the same number of followers at any point of time which they initially purchased. To assess whether this is true or not, we continuously monitored the followers gained by our dummy customer accounts for the premium market. We observed that only 3 merchants provided us lesser followers than promised and rest 54 merchants provided us either the exact number of requested followers or more (as previously seen in Table III). However, after the date of purchase and the gain of requested amount of followers, we observed constant drop in the follower count in case of all the customers. We further investigated the drop in follower count by taking hourly snapshots of each of the customer profile and observed constant fluctuations in the follower count during each day. Figure 3 shows this phenomena for one of the popular premium merchants in our dataset.

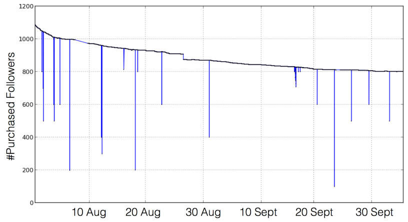


Fig. 3: Shows the dips in follower count every hour for one of the most prominent premium merchants in our dataset. Few minutes/hours after the dips, the follower count rises back.

We also observed that there were several dips in the follower counts in both markets, however, we did not gain new users as followers. Users from the same set of initially obtained users kept unfollowing and following us back. We posit that the reason behind such following behaviour is that since the follower accounts' activities are controlled by the merchants, they optimize who to follow to cater to a larger customer base without raising suspicion of following too many users at a given point of time.

TABLE V: List of most popular promises and guarantees made by the merchant to customers.

	Promises by Merchants to Customers
Freemium	*60+ new followers per ride* *promotional status updates* *never alter profile information*
Premium	*new followers every minute* *Ad free* – No promotional tweets *3 month no drop guarantee* – follower retention policy *genuine profile* – legitimate users will follow *delivered in 1-2 days*

2) Quality evaluation of merchants: Catering to the needs of customer is the key to successful busineses. Twitter follower merchants make several promises to attract and retain customers, as previously indicated in previous section. Therefore, we calculated QoS of the merchants to understand how well they are catering to their customers. We used the following definition of QoS and its parameters as described earlier in Equation (1)

$$\textbf{QoS} = q(PERFORM, EXPECT, DISCONFIRM)$$

For simplicity, we gave equal weightage to all the promises, and hence $p_k = 1$ in Equation (2). DISCONFIRM is the discrepency between performance of merchant and expecation of the customer for a particular promise made by the merchant. For a specific merchant, we calculated the collective QoS based on all the promises $\{SA_1...SA_N\}$ which it provides as -

$$QoS = \frac{\sum_{i=1}^{N}\left[1 - \left(\frac{EXPECT_i - PERFORM_i}{PERFORM_i}\right)\right]}{N}$$

Note that in case a merchant overdelivers a certain promise, then the above formulization of QoS for that specific promise gets a value > 1, hence rewarding the merchant. This gave us a normalized value of QoS for all the merchants. Figure 4 shows the QoS curve for both freemium and premium merchants. The knee point of the QoS curve for freemium market lies at X=0.1, Y=0.3; this indicates that 90% freemium merchants have a QoS value of 0.3 or less. The knee point for premium merchants is at X=0.05, Y=0.28 indicating that 95% premium merchants in our dataset have a normalized QoS value of 0.28 or lesser. The highest QoS for freemium market is 0.82, whereas for premium we found it to be 0.78. This shows that overall, QoS for freemium market is higher than that of freemium market. We further investigated and found that the violation of *no drop guarantee* is prime reason behind the low QoS for premium market. We had earlier seen this phenomena of follower drop in Figure 3. Frequent drops in follower can raise a red flag against the customer, and hence the merchants which deliver followers exhibiting such phenomena were penalized in our formulization of QoS.

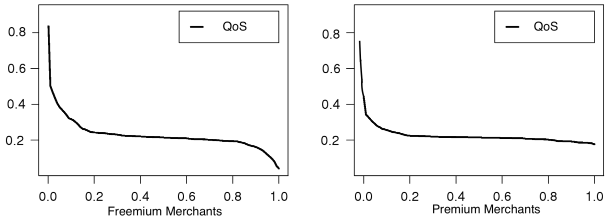


Fig. 4: Quality of Service of Freemium and Premium Markets.

3) Few merchants are market leaders: Now we evaluate which merchants attract the highest share of user base. In order to do so, we computed the popularity of each merchant by using two metrics - Alexa ranking of merchant website, and number of promotional tweets. [1]

Alexa Ranking: Alexa rank measures website's popularity based on the traffic to that website. For each merchant, we extracted the global rank of its website and then computed the normalized Alexa rank for merchant M_i as followed –

$$Alexa\_Norm_i = 1 - \frac{AlexaRank_i}{\max\limits_{i \in \{1...N\}}(AlexaRank_i)}$$

where N is the total number of merchants. This gives us a normalized measure of website traffic between 0 and 1.

Social Media Popularity: We also used social media popularity of the merchant website by collecting the promotional tweets advertising the merchant if any. We searched for each merchant's URL and its Bitly shortened version using Twitter search API. We then defined the OSN popularity for each of the N merchants as –

$$OSN\_Popularity_i = \frac{NumTweet_i}{\max\limits_{i \in \{1...N\}}(NumTweet_i)}$$

where $NumTweet_i$ is the number of promotional tweet for merchant M_i. Using these two metrics, we finally calculated the overall popularity (*Popularity Score*) of a merchant website by taking an average of normalized Alexa ranking and OSN Popularity. Figure 5 shows the distribution of popularity of all the merchant websites. We noticed that 5 of the merchants had very high popularity as compared to other merchant websites. The top 5 merchants had a normalized popularity score of more than 0.71 (71%). This indicates that there exist an *oligopoly* hierarchical structure amongst the merchants. [2] That is, there are few market leaders which attract most of the audience. We

[1] Alexa – http://www.alexa.com/

[2] Oligopoly: http://www.economicsonline.co.uk/Business_economics/Oligopoly.html

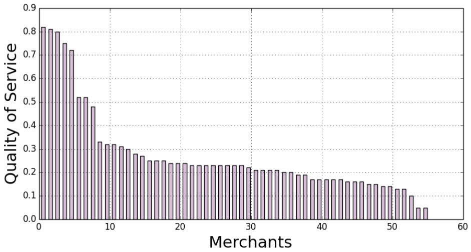


Fig. 5: Quality of Service of Freemium and Premium Markets.

notice that 4 of the market leaders have Twitter verified customers in our dataset. [3] The data collection from freemium merchant websites show that atleast 10 Twitter verified accounts subscribed to services of one of the top 4 merchants represented in Figure 5. This also indicates that market leaders attract high profile customers.

B. Underground Market Customers

1) Spammers, wannabes and celebrities: To understand who are the customers of follower market, we used our dataset of 171,234 customers collected from the merchant websites. We found out whether they are listed, verified and analyse their *bio*. Figure 6 shows that many customers use *'follow'*, *'artist'*, *'director'*, *'music'* in their bio. This indicates that these users are probably *wannabe* artists and are trying to attract a large following. We also found that 10% of the users had posted atleast one or more URL blacklisted by Google Safebrowsing [4] or Phishtank. [5] We also noticed that 10 of the customers we acquired in our dataset were Twitter verified users and had posted atleast one promotional tweet (which was soon deleted) about the merchant website. This shows that Twitter verified accounts, i.e., celebrity accounts, also use freemium merchants to boost their follower count.

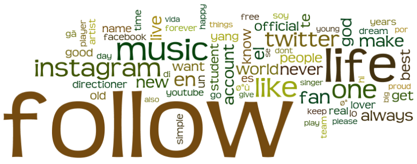


Fig. 6: Wordcloud of bio of the customer profiles.

2) Market leaders attract prominent customers: To identify the prominent customers, we used 'Klout' [6] score. This is a popular tool to measure influence based on various factors like followers, freinds, retweets and favourites. The average Klout score for the social media users is 40. [7] We found that 30% customers had a Klout score of more than 40. Out of these prominent users, 81.7% users subscribed to atleast one merchant with a *Popularity Score* greater than 0.72. This indicates that the merchants which are more popular and market leaders attract prominent customers who have a higher reputation.

C. Phony Followers of Underground Market

In this section, we present an anatomy of the phony followers which we recieved from various merchants. We find key identifiers which can be helpful to distinguish from legitimate users and hence help us to build an effective detection model.

1) Phony followers have low social engagement: We explored how the purchased follower accounts are connected with their friends. We measured social engagement of users with their friends in form of retweets, @-mentions and favorite count.

a) RTs and @-mentions: We observed that a large fraction of purchased accounts post only retweets instead of original content. We further explored whether these users retweet the content of their friends or not. If RT_{count_i} is the number of tweets the user has retweeted of her friend u_i and she has N friends, then

$$RetweetRatio = \frac{\frac{RT_{count_i}}{\Sigma_{i=1}^{N} RT_{count_i}}}{N * RT_{total}}$$

RT_{total} is the total number of retweets done by the user. This *Retweet Ratio* quantifies the number of friends a user has retweeted and the number of times she retweeted them.

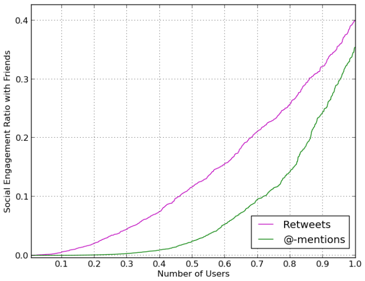


Fig. 7: Social engagement of Purchased Users with their Friends.

Similarly we define the @-mention ratio to determine whether the user engages in conversations with her friends and to what extent.

$$@mentionRatio = \frac{\frac{@_{count_i}}{\Sigma_{i=1}^{N} @_{count_i}}}{N * @_{total}}$$

where $@_{total}$ is the total number of @-mentions by the user. We observed in Figure 7 that the highest Retweet Ratio score is 0.45 and the @-mention ratio is 0.35. This shows that though a large fraction of purchased accounts post only retweets, its not the tweets of their friends which they are retweeting. Similarly, low @-mention ratio suggests that purchased followers do not mention their friends. We found the maximum @-mention ratio with the followers of purchased users to be 0.32. This indicates that purchased followers are low quality users and do not engage in conversations with their friends or followers.

b) Language overlap with Friends and Followers: We obtained the language of every tweet as returned by Twitter API. Figure 8 shows the distribution of language of purchased accounts. We observed that a 52% of the users tweet in spanish. We also found that the purchased followers tweet and retweet in multiple languages as shown in Figure 8. Thirteen percent users used 5 or more languages. Only 32% users posted tweets in less than or equal to two languages. We next found the overlap of language amongst the purchased accounts with their followers and friends. Users tweet and retweet in multiple languages. We calculated the *Language Overlap Score* for each user defined as

$$LangOverlap = \frac{\Sigma_{i=0}^{N} overlap_i}{N}$$

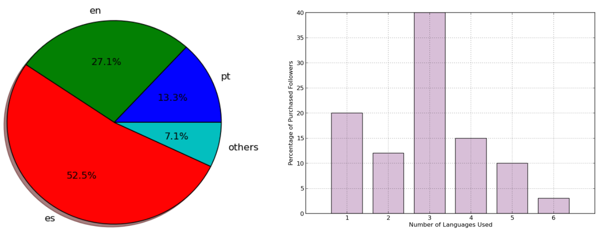
[3] Twitter verified users https://support.twitter.com/articles/119135-faqs-about-verified-accounts

[4] Google Safebrowsing https://developers.google.com/safe-browsing/

[5] Phishtank www.phishtank.com

[6] http://www.klout.com

[7] http://support.klout.com/customer/portal/articles/679109-what-is-the-average-klout-score



(a) Distribution (b) Number of Languages

Fig. 8: Languages used by Purchased Followers.

where N is the total number of friends or followers. If L_f is the set of languages used by the friend/followers and L_u is the set of languages used by the purchased user then $overlap_i$ with each friend/follower u_i is defined as $overlap_i = 1$ if $|L_f \cap L_u| \neq 0$, else 0. We used the *Language Overlap* score to determine how many users tweet in same language as their friends or followers. Figure 9 shows that 80% users had an Overlap Score = 0.37 with their followers and Overlap Score = 0.68 with their friends. This indicates that a large fraction of purchased follower accounts do not care about the content posted by the users they are following. Also, the followers of these users do not have a high language overlap with them.

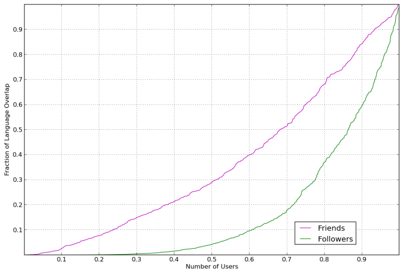


Fig. 9: Language Overlap of Purchased Followers with their Friends and Followers.

2) Phony followers have low social reputation:

a) Follower-Friends Ratio: We now look at the relationship between amount of followers and friends for purchased follower accounts. On Twitter, 'followers' of a person are the users which subscribe to the posts of that person, i.e., who 'follow' her. The 'friends' of a person are the users whom she subscribes to. The average number of followers per existing account is 68 and the average number of friends is 60 on Twitter. Figure 10 shows that the follower/friends ratio fits the power law ($\alpha = 1.8209$, *error* $\sigma = 0.029$). We observed that 94% purchased followers have the follower/friends ratio as only 0.1 and none of the purchased followers had more followers than friends. Low follower/friends ratio indicates that the user does not have a good following, therefore indicating a low social importance.

b) Klout Score: To measure the social influence, we used Klout score. We recall that the average Klout score for the social media users is 40. However, as shown in Figure 11, we found that 90% of the purchased followers had a Klout score of less than 20. This shows that these accounts do not involve in discussions with other users and have a low influence score.

3) Phony followers exhibit high unfollow entropy: We found that the purchased follower unfollowed a large number of users regularly. To quantify this behaviour, we calculated the *unfollow entropy* of all the purchased followers. We observed each purchased follower over a span of 15 days and collected her hourly followers. We define normalized *unfollower entropy* H for a user u_n as the following

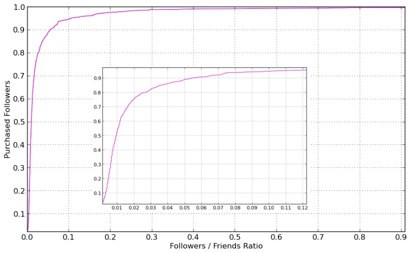


Fig. 10: Follower-Friend ratio of purchased follower accounts.

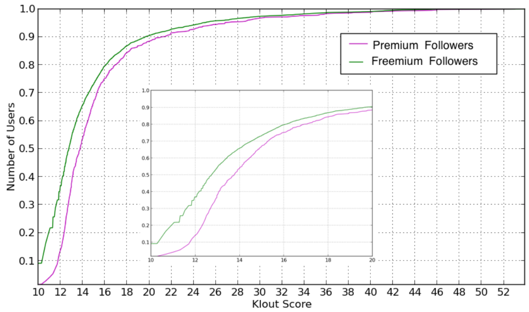


Fig. 11: CDF of Klout Score of Purchased Followers.

$$H_{u_n} = -\frac{\Sigma_{i=1}^{T} p_n(f_i) log(p_n(f_i))}{N}$$

where, $p_n(f_i)$ is the probability that the user u_n will unfollow at time t_i. The probability function is defined as

$$p_n(f_i) = \frac{ucount_i}{\Sigma_{i=1}^{T} ucount_i}$$

where T is the number of days for which we monitor the purchased follower and $ucount_i$ is the number of users she unfollowed on i^{th} day. A higher value of unfollow entropy signifies that the user exhibits a suspicious unfollow pattern. Figure 12 shows that a large fraction of purchased followers had a high unfollow entropy. The normalized entropy rate for 23% purchased followers was as high as 0.76 and only 8% users had a normalized unfollow entropy less than 0.21. To find out whether the users with higher unfollow entropy have lower quality than other users, we compared their normalized unfollow entropy rate with *Klout* score. We found a strong negative correlation (*Pearson correlation coefficient = -0.73*) indicating that users with higher unfollow entropy rate have low social reputation.

VI. Prediction of Suspicious Following Behaviour

In the second part of our study, we built a supervised predictive model to detect suspicious following behaviour on Twitter. In this section, we explain the feature set used for the classification task and the experimental setup.

A. Features for Classification

For our prediction task to detect suspicious following behaviour, we explored *user profile*, *network*, *content* and *user behaviour* based features. In all, we explore 18 features for our classification task as described in Table VI. *User profile* based features focus upon properties of the Twitter user profile information like 'bio' and 'profile URL'. The *network* based features describe the relationship of the user with her friends and followers. We next explored the *content* based

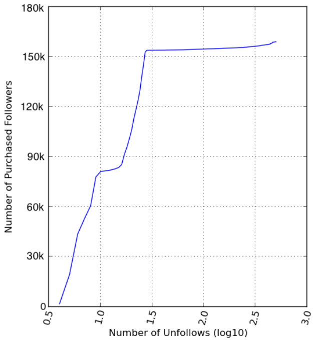


Fig. 12: Unfollow Entropy Rate for Purchased Followers. A large number of followers in our data follow-unfollow their friends multiple times.

features to understand the nature of tweets posted by the user and also investigatde the *behavioural* features to understand the tweeting patterns and follow dynamics exhibited by the user. For the *network* based features, we constrained our analysis to single hop network of the users due to Twitter API rate limit restrictions. Also, we kept our *content* based analysis limited to stylistic features of tweets due to the presence of multilingual users in our dataset and the complexity of computation due to transliterated text, misspellings and use of short hand language. Table VI enlists all the feature sets we used for our prediction task.

TABLE VI: Description of the feature sets used for prediction of users with suspicious following behaviour.

Set	Category	Features
A	User Profile	presence of bio presence of URL in bio number of posts social reputation
B	Network	follower / friends ratio number of followers
C	Content	hashtags per tweets spam words used per tweet length of tweet number of languages used number of RTs per tweet @mentions per tweet
D	Behaviour	unfollow entropy rate RT engagement score @mention engagement score language overlap time since last tweet tweets per day

We explained some of these features in the previous section; here we describe how we calculated the values of remaining features:

*a) **Presence of bio and URL**:* Some Twitter users give description about themselves on their profile which is called *bio*. We checked the presence of *bio* for each user under inspection. We also checked whether the user has mentioned any external URL in her *bio* and use this as a feature.

*b) **Hashtags per tweet**:* We calculatde the average number of hashtags used per tweet. We define this metric as

$$hashtag/tweet = \frac{\Sigma_{tweet=0}^{N} \#hashtags}{\#tweets}$$

*c) **Spam words used per tweet**:* In the earlier section, we noticed that a fraction of purchased follower accounts also spread spam and malicious content. To detect spam in the tweet content, we used a spam word lookup list [8] and define the following metric

$$spamwordspertweet = \frac{\Sigma_{tweet=0}^{N} \#spamwords}{\#tweets}$$

*d) **Time since last tweet**:* We found that purchased followers exhibiting suspicious following behaviour have very less tweeting activity and are often inactive. To measure time since the account has been inactive, we found the difference in time in seconds since the latest tweet with the time of our experiment.

B. Experimental Setup and Classification

For our classification experiment, we considered the 170k public purchased followers as our true positive dataset of suspicious follow behaviour. For the negative class (legitimate follow behaviour), we picked random 170k users from Twitter stream using the streaming API. However, a balanced dataset as ours may create a sample bias. Therefore, to ensure valid results and eliminate the bias, we undersampled our negative class. We drew 10 random but independent subsets from the set of 170k legitimate users (-ve class) and trained 10 classifier models based on these 10 subsets along with the 170k samples of the suspicious follow behaviour users (+ve class). We then used 10 fold cross validation and reported the average results for our prediction task. We treat the detection of suspicious follow behaviour as a two class classification problem. In order to detect such behaviour, we used several supervised learning algorithms like Naive Bayes, Gradient Decent, Random Forest etc. However, we achieved highest accuracy and overall best results with *Support Vector Machine* (SVM). We used a non-linear SVM with the Radial Basis Function (RBF) kernel for our experiment. Table VII gives the details of our experimental setup - dataset description and the parameter values for the SVM classification algorithm. In order to assess the effectiveness of features, we repeated the classification experiment by incrementally adding each feature set. For evaluation, we used 70-30 split of the training and testing dataset. We used 10 fold cross validation to report our results.

TABLE VII: Description of the experimental setup for prediction fo suspicious following behaviour.

Dataset	342,000 users
Suspicious (+ve class)	170,000 users
Legitimate (-ve class)	170,000 users (10 times)
Classifier	SVM
C	1,000
alpha	20.0
Classification Runs	10
Feature Sets	{A}, {A, B}, {A, B, C}, {A, B, C, D}
Train-Test Split	70%-30%
Cross Validation	10-fold

C. Classification Results and Evaluation

Table VIII shows the confusion matrix for our classification task. The confusion matrix defines the percentage of false negatives and false positives. We were able to accurately classify 88.5% users with suspicious follow behaviour and 89.9% users with legitimate behaviour. This shows that we are able to detect suspicious following behaviour to a good extent. For the evaluation of our classification result, we used the standard evaluation metrics – accuracy, F-measure and AUC.

[8] http://www.mailup.com/spam-words-to-avoid.htm

TABLE VIII: Confusion Matrix – Classification Results of distinguishing legitimate users from those exhibiting suspicious following behaviour.

		Predicted	
		Suspicious	Legitimate
True	Suspicious	88.5	11.5
	Legitimate	9.7	89.9

As discussed in the previous section, we incrementally added feature sets to evaluate the effectiveness of all the features. Figure 13 shows the performance of our classifier on Accuracy, F1 score and AUC metrics when feature sets are incrementally added. We noticed that each feature set has a positive effect on the performance of the classifier across all metrics. We also observed that adding behavioural based features suddenly increase the overall accuracy of our classification model. We received a maximum accuracy of 89.2%.

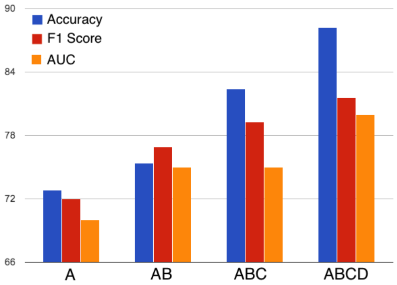


Fig. 13: Classification accuracy to predict suspicious following behaviour on incremental feature addition.

D. Feature Importance

We found that behavioural features are important to detect suspicious follow behaviour. *Unfollow entropy* rate played an important role. Some of the most informative features we received after our classification task were *unfollow entropy, RT-engagement ratio, @mention-engagement ratio, Language Overlap and Social Reputation*. The other informative and discriminative features were the use of multiple hashtags and spam words in the tweets. The profile based features were the least helpful in detection of suspicious follow behaviour. One possible reason for this could be that a large fraction of legitimate users do not add a *bio* or engage in heavy conversations on Twitter.

VII. Conclusion

In this study, we present a landscape of Twitter underground follower market. We focus on the building blocks of follower market - *merchants*, *customers*, *phony followers*. We used a dataset of 60 freemium and 57 premium merchants for our study which were most responsive and active at the time of our experiment. We collected 170k phony followers from these merchants. Though there exist millions of compromised and synthetic accounts controlled by the follower merchants, we study a random and large portion of underground market over a course of 4 months. Therefore, we posit that our results would be scalable over a much larger network of underground merchants.

To summarise, we present the following in this study (i) We measure the QoS and uncover the underlying hierarchy of merchants. We discover an *oligopoly* structure of the merchants, (ii) We analyze the reputation and profile attributes of the market customers to understand who they are and which merchants they subscribe to, (iii) We study suspicious following behaviour of the phony followers and build a supervised learning model to distinguish them from legitimate users. This is the first study to landscape all aspects of an underground market to understand its underlying structure and characteristics.

References

[1] The dirty business of buying instagram followers. http://racked.com/archives/2014/09/11/buy-instagram-followers-bloggers.php, 2014.

[2] Fake twitter followers generate millions of dollars. http://cir.ca/news/fake-twitter-followers, 2014.

[3] Review site yelp battles extortion claims. http://www.washingtonpost.com/business/review-site-yelp-battles-against-extortion-claims/2014/10/13/502a6ff4-5298-11e4-b86d-184ac281388d_story.html, 2014.

[4] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos. Copycatch: stopping group attacks by spotting lockstep behavior in social networks. In *Proceedings of the 22nd international conference on World Wide Web*, pages 119–130. International World Wide Web Conferences Steering Committee, 2013.

[5] R. N. Bolton and J. H. Drew. A multistage model of customers' assessments of service quality and value. *Journal of consumer research*, pages 375–384, 1991.

[6] J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring pay-per-install: The commoditization of malware distribution. In *USENIX Security Symposium*, 2011.

[7] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 2010.

[8] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang. Detecting suspicious following behavior in multimillion-node social networks. In *WWW*, 2014.

[9] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Félegyházi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, et al. Click trajectories: End-to-end analysis of the spam value chain. In *Security and Privacy (SP)*, 2011.

[10] M. Motoyama, K. Levchenko, C. Kanich, D. McCoy, G. M. Voelker, and S. Savage. Re: Captchas-understanding captcha-solving services in an economic context. In *USENIX Security Symposium*, 2010.

[11] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker. Dirty jobs: The role of freelance labor in web service abuse. In *USENIX Security Symposium*, 2011.

[12] G. Stringhini, M. Egele, C. Kruegel, and G. Vigna. Poultry markets: on the underground economy of twitter followers. In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*, pages 1–6. ACM, 2012.

[13] G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and B. Y. Zhao. Follow the green: growth and dynamics in twitter follower markets. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 163–176. ACM, 2013.

[14] K. Thomas, C. Grier, D. Song, and V. Paxson. Suspended accounts in retrospect: an analysis of twitter spam. In *IMC*, 2011.

[15] K. Thomas, D. Iatskiv, E. Bursztein, T. Pietraszek, C. Grier, and D. McCoy. Dialing back abuse on phone verified accounts. 2014.

[16] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson. Trafficking fraudulent accounts: The role of the underground market in twitter spam and abuse. In *USENIX Security*, pages 195–210. Citeseer, 2013.

[17] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, M. India, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In *Proceedings of the 23rd USENIX Security Symposium (USENIX Security)*, 2014.

[18] G. Wang, C. Wilson, X. Zhao, Y. Zhu, M. Mohanlal, H. Zheng, and B. Y. Zhao. Serf and turf: crowdturfing for fun and profit. In *WWW*, 2012.