The Many Faces of Link Fraud

Neil Shah*, Hemank Lamba*, Alex Beutel†, Christos Faloutsos*
Carnegie Mellon University*, Google Research†
Email: {nshah, hlamba, christos}@cs.cmu.edu, alexbeutel@google.com

***Abstract*—Most past work on social network link fraud detection tries to separate genuine users from fraudsters, implicitly assuming that there is only one type of fraudulent behavior. But is this assumption true? And, in either case, what are the characteristics of such fraudulent behaviors? In this work, we set up *honeypots*, (“dummy” social network accounts), and buy fake followers (after careful IRB approval). We report the signs of such behaviors including oddities in local network connectivity, account attributes, and similarities and differences across fraud providers. Most valuably, we discover and characterize *several* types of fraud behaviors. We discuss how to leverage our insights in practice by engineering strongly performing entropy-based features and demonstrating high classification accuracy. Our contributions are (a) *observations*: we analyze our honeypot fraudster ecosystem and give surprising insights into the multifaceted behaviors of these fraudster types, and (b) *features*: we propose novel features that give strong (>*0.95* precision/recall) discriminative power on ground-truth Twitter data.**

I. Introduction

What are the characteristics of fraudulent accounts in online social networks? Understanding the behavior and actions of fraudsters and incorporating these insights into detection algorithms is paramount to preventing fraud. Previous works in social network fraud detection have primarily focused on leveraging signature properties of fraudsters for detection including temporally synchronized behavior [1], excessively dense [2] and oddly distributed [3] graph connectivity, uncommon account names [4] and spammy links [5]. Our work instead focuses on establishing the veracity and applicability of these assumptions. We ask: do all fraudsters behave similarly, or are there multiple signatures? Since fraud detection is an adversarial setting in which fraudsters are constantly adapting, it is important to constantly monitor and evaluate the strategies that they employ to profitably perform ingenuine actions to better inform future detection mechanisms. We focus on one particular setting of social network fraud called *link fraud* which involves the use of fake, *sockpuppet* accounts to create links (graph connections) which represent followership. Fake links artificially inflate the follower count of customers, deceive authentic users and hinder the performance of machine learning algorithms which rely on authentic input to recommend relevant content. Our informal goal is as follows:

Problem 0 (Informal)**.** ***Given*** *a social network,* ***identify*** *patterns of link-fraudster behavior, and* ***extract features*** *which can* ***discriminate*** *fraudsters from genuine users.*

To study the behavior of fraudsters, we employ the use of *honeypots*, or dummy accounts on which we solicit fake Twitter followers sourced from various providers. Honeypots give us a clear signal of fake activity untainted by authentic behavior. Upon setting up the honeypot accounts and purchasing fake followers, we collect a rich representation of the fraudster ecosystem which we subsequently analyze. We discover a number of *different* patterns of fraudster behaviors, with the possibility of even more types (see Figure 1). Furthermore, we study and characterize the network connectivity properties and attribute distributions which are exhibited by these different types of fake followers, showing the suspicious patterns that each of them induces. Our work shows that assumptions made in previous detection works with respect to unimodality of fraud are not necessarily justified. Summarily, this work makes and explores the following key insight:

Key Insight (Fraud Multimodality)**.** *There are multiple types of link fraud which exhibit different network connectivity and account attributes.*

We offer the following major contributions:

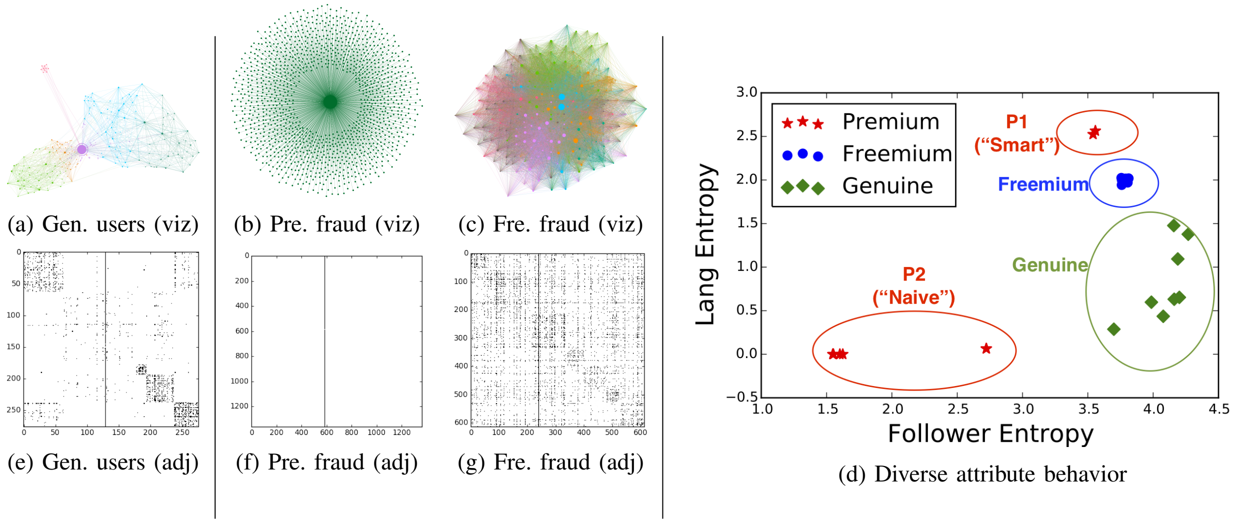
- **Observations**: We collect and analyze ground-truth fraudster data, observe and make numerous insights about the *multiple* behaviors of link fraudsters.
- **Features**: We carefully engineer novel entropy-based features which allow us to discern fraudsters from genuine users with nearly perfect F1-score.

II. Related Work

Prior works have shown the use of fake accounts for social media followers [6], email accounts [7], Facebook likes [1], etc. These accounts are often used to spread spam [8] and misinformation [9]. [10] estimates that the fake follower market produces $360 million per year. [11] explores underground markets providing fake content, reviews and solutions to security mechanisms. [6] studies several fraud providers over time and describes trends in pricing, account names and IP diversity. [12] compares growth rates of accounts with legitimate and fraudulent followers. [13] observes the varying retention and reliability of various fraud providers. Comparatively, our work is the first to identify major social graph differences between fraud types and across providers, and propose novel entropy-based features for capturing these behaviors. Most previous fraud detection works focus on feature-based classification [14], [15], [16], subgraph analysis [1], [17] and spectral methods [3], [18], [2].

III. Know Thy Enemy: Characterizing Link Fraud

In this section, we describe our data collection process and introduce relevant metrics for illustrating numerous insights



(a) Gen. users (viz) (b) Pre. fraud (viz) (c) Fre. fraud (viz)

(e) Gen. users (adj) (f) Pre. fraud (adj) (g) Fre. fraud (adj)

(d) Diverse attribute behavior

Fig. 1: **Freemium (Fre) and premium (Pre) fraud types have different local network structure and account attributes compared to genuine behavior.** Nodes are colored by modularity class, and sized proportional to in-degree in (a)-(c). The associated, reordered adjacency matrices are shown in (e)-(g) – the vertical line in each spyplot indicates the the central node. Notice the block community structure in genuine followers compared to the star structure for premium and near-clique structure for freemium followers. (d) shows differences in attribute (language and follower) entropy over the various behaviors, showing how fraud patterns skew attribute distributions away from genuine ones.

about fraudster network connectivity and account attributes.

A. *Preliminaries*

TABLE I: Honeypot account summaries.

Service	Type	Cost	Followers bought	Followers delivered	Followers remaining
fastfollowerz	Premium	$19	1000	1060 1060	1059 1059
intertwitter	Premium	$14	1000	1099 1102	977 974
devumi	Premium	$19	1000	1360 1354	1358 1354
twitterboost	Premium	$12	1000	1361 1350	1361 1350
plusfollower	Freemium	£9.99	1000	1094 1078	748 737
hitfollow	Freemium	£9.99	1000	926 937	623 638
newfollow	Freemium	£9.99	1000	884 883	600 589
bigfolo	Freemium	£9.99	1000	872 865	594 577

1) Purchasing Fake Followers: We solicit fake followers from popular services with high rank on Google search results using queries such as "buy Twitter followers," as these services likely serve the majority of customers.

In surveying these, we notice there are two prevalent models of service – we call these *premium* and *freemium*. Premium services offer customers followers in tiers (1K, 5K, 10K, etc.) for payment and ask only for the customer's Twitter username. Freemium services offer a paid option as in premium services, but also offer a free alternative which requires the user to provide their Twitter *login details* to the service – in return for these details, the services promise to direct a small number of followers to the account.

We next setup a pool of honeypots by repeating the Twitter account creation process using monikers from online screen-name generators. Account creation was spread over IPs due to Twitter's IP-based account creation limits. We enrolled in the basic, 1K follower packages from 8 services (4 freemium, 4 premium) using 2 honeypots per service. We used 2 accounts per service in order to examine account reuse policies for individual providers. Table I summarizes the details of this process. Honeypots were created on the same day, and follower purchases were done simultaneously. Furthermore, the honeypots attracted no followers themselves prior to our purchases. Thus, we posit that all our honeypot followers are fake.

2) Data Collection: We use the rate-limited REST API to scrape data relevant to our operation from Twitter. Prior to purchasing followers, we start several Python scripts which poll the API at varying intervals due to rate-limits, and populate a database with the information. We collect honeypot account details, their followers' IDs and details, and the IDs and details of the followers' own friends and followers to scope out other accounts with known fake links. These details include Tweet count, language, profile descriptions and settings, etc. Our carefully engineered tracking scripts are made public at `https://goo.gl/5wnYpc`.

3) Summary Statistics: We introduce several data structures. An *ego network* (egonet) consists of a central node (ego) and the neighboring nodes and edges between them. In this paper, we examine per-service egonets, taken as the union of the individual egonets of both honeypots per service. This allows us to study how fraud providers reuse their accounts. We also propose and define the *boomerang network* as the per-service egonet plus the out-links of the followers – the

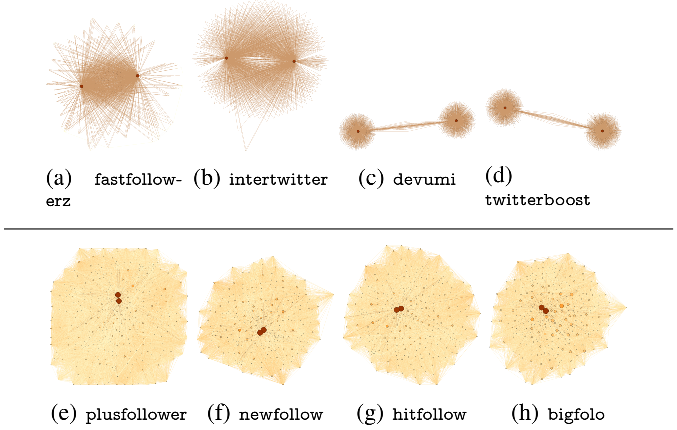


Fig. 2: **Premium fraudsters (top) form overlapping stars whereas freemium ones (bottom) form dense, near-cliques.** Subplots show per-service egonets with honeypots in dark-red – darker color and larger size indicates higher in-degree.

structure goes "1 step back, 1 step forward" from the honeypot. The boomerang network focuses on the external relationships formed by fake followers, whereas the egonet focuses on the internal relationships between fake followers and honeypots.

We additionally define several metrics for convenience. Given nodes $\mathcal{N}$ and edges $\mathcal{E}$, *density* is defined as $|\mathcal{E}|/(\mathcal{N}(\mathcal{N}-1))$, or the proportion of existing to possible edges. Given node sets $\mathcal{N}_1$ and $\mathcal{N}_2$ and the edge set between them $\mathcal{E}_{\mathcal{N}_1 \to \mathcal{N}_2}$, *bipartite density* is defined as $\mathcal{E}_{\mathcal{N}_1 \to \mathcal{N}_2}/(|\mathcal{N}_1||\mathcal{N}_2|)$, or the proportion of existing to possible edges between the parts. *Overlap coefficient* is defined as $|\mathcal{N}_1 \cap \mathcal{N}_2|/min(|\mathcal{N}_1|, |\mathcal{N}_2|)$ and represents the degree of overlap between sets. The *multiple systems estimate* (MSE), defined as $(|\mathcal{N}_1||\mathcal{N}_2|)/|\mathcal{N}_1 \cap \mathcal{N}_2|$ estimates population size based on overlap between random samples. Given the graph has T triangles and W wedges (connected triples), *transitivity* is defined as $3T/W$ and denotes the degree of triadic closure. Given the graph's bidirectional edge set $\mathcal{E}_{\leftrightarrow}$ and edge set $\mathcal{E}$, *reciprocity* is defined as $\mathcal{E}_{\leftrightarrow}/\mathcal{E}$ and indicates reciprocal frequency. Lastly, given distribution X with outcomes $(x_1 \dots x_n)$, *entropy* is defined as $-\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$ and measures information content in bits.

B. Network Observations

Link fraud impacts graph structure by its mission constraints. But how? In this section, we leverage network analysis tools on the aforementioned induced subgraphs to characterize effects of fraud on surrounding network structure, and compare and contrast premium and freemium fraud.

1) Ego Network Patterns: Figure 2 shows the per-service egonets for each of the 8 providers. The honeypots (egos) are the two large and dark orange colored nodes in each subfigure. Our analysis reveals notable differences in egonet structure between freemium and premium providers. We see that the premium egonets (first row) have a star/bipartite structure: each honeypot is the hub of a star, and satellite nodes overlap and are disconnected. Conversely, freemium egonets have denser, near-clique structure which suggests heavy connectivity between neighboring nodes.

TABLE II: Egonet summary statistics.

	Service	# Nodes	# Edges	Density	Transitivity	Reciprocity
Premium	fastfollowerz	1,066	2,289	.002	.001	.000
	intertwitter	1,051	2,003	.002	.00006	.000
	devumi	2,681	2,712	.0003	.000	.000
	twitterboost	2,680	2,711	.0004	.000	.000
Freemium	plusfollower	920	51,868	.061	.288	.411
	newfollow	755	37,052	.065	.294	.408
	hitfollow	782	41,879	.068	.305	.416
	bigfolo	749	36,043	.064	.294	.413

Egonet statistics in Table II further lend credence to the visual differences we observe from Figure 2, giving us the following insight:

Insight 1 (Egonet Sparsity)**.** *Premium fake followers rarely follow each other, resulting in sparse egonet structure. Freemium fake followers have dense egonet structure.*

As shown in Table II, freemium providers are an order of magnitude denser than the densest premium ones – all 4 providers have 6-7% density. Of these providers, fastfollowerz and intertwitter have significantly higher density and transitivity than devumi and twitterboost. These providers also have no reciprocity, indicating one-way relationships. Contrary to premium services, freemium providers have high reciprocity of 40-42% suggesting frequent "follow-backs."

These differences are likely because freemium services accumulate user pools and trade follows amongst the free users. These accounts create a dense subgraph, and are also used by providers to serve paid customers and turn a profit. Comparatively, premium providers are unable to use free users' accounts and must create fake ones.

As we expect fraudsters to act in a manner that maximizes profit, *what motivates these differences between freemium and premium providers?* We propose a rationale: If we consider that each account has a budget of edges it can create without being suspended, it seems that premium providers greatly underutilize accounts compared to freemium ones. This is because for fraudsters, delivering more links while avoiding suspension is strictly better as it means that they can either serve more customers or artificially inflate their own popularity. It turns out that premium providers are actually able to *better* utilize followers than freemium ones by capitalizing on external versus internal connectivity, as we see next.

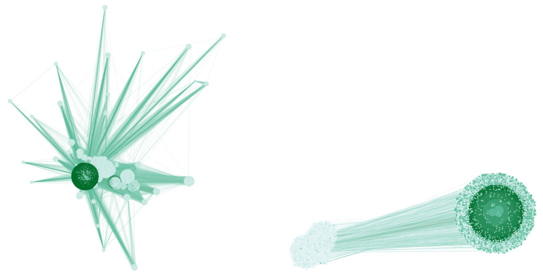
2) Boomerang Network Patterns: Figure 3 shows 2 boomerang networks, one for bigfolo and twitterboost, each representative of a different fraud strategy. Visually close nodes have similar connectivity. As with egonets, we again see a large contrast in the structures of these two providers. Figure 3a shows the dense internal connectivity of bigfolo's freemium followers, in conjunction with the sparser and distributed external links to friends. Conversely, Figure 2d shows sparse internal linkage between twitterboost's premium followers on the left, but dense near-bipartite external linkage to cus-

TABLE III: Boomerang network summary statistics.

	Service	# Nodes	# Edges	Bip. Density
Premium	fastfollowerz	40,486	491,458	.012
	intertwitter	176,921	2,383,251	.013
	devumi	67,893	2,495,586	.014
	twitterboost	68,297	2,474,759	.014
Freemium	plusfollower	646,901	1,352,253	.002
	newfollow	616,824	1,221,574	.003
	hitfollow	558,100	1,172,248	.003
	bigfolo	574,823	1,157,672	.003

TABLE IV: Fraud providers have varying account reuse habits.

	Service	# Nodes	Overlap	Est. Pool # Nodes
Premium	fastfollowerz	1,064	.996	1,064
	intertwitter	1,049	.953	1,051
	devumi	2,679	.024	55,719
	twitterboost	26,78	.024	55,677
Freemium	plusfollower	918	.815	954
	newfollow	753	.765	798
	hitfollow	780	.802	814
	bigfolo	747	.774	791



(a) bigfolo (fre.) (b) twitterboost (pre.)

Fig. 3: **Freemium followers have dense internal and sparse external connectivity (left), and vice versa for premium followers (right).** Subplots show boomerang networks, with darker node color and larger size indicating higher in-degree.

tomers (including honeypots) on the right. Table III's further substantiates the following:

Insight 2 (Boomerang Density)**.** *Premium fake followers are frequently reused to follow customers, resulting in dense external connectivity in the boomerang network. Freemium fake followers are less reused to follow customers, and hence have sparse external connectivity.*

Note that the relative values for density, node and edge count in Table III are inverted compared to Table II between freemium and premium providers.Premium providers' bipartite density indicates existence of nearly 1-2% (a huge amount) of all possible edges between fake followers and their combined friends, while freemium providers have an order of magnitude lower bipartite density than premium ones. Node to edge ratios are also much higher for premium providers – fastfollowerz and intertwitter are 1:14, and devumi and twitterboost are roughly 1:37 compared to only 1:2 for the freemium providers. Additionally, freemium boomerang networks have higher node count than premium ones. Intuitively, since freemium followers are real accounts, they have an expansive set of (real) friends, whereas premium followers are synthetic and have a smaller set of friends/customers.

3) Network Overlap Patterns: Next, we ask: how extensively do services reuse accounts? Do they do so in the same ways? Furthermore, is there any overlap across providers?

a) Intra-Network Patterns: In Table IV, we present overlap coefficients between followers per service. Assuming that the followers are randomly sampled from the service's pool, we also compute an estimated total pool size per provider using the MSE statistic. The overlap and pool size estimates suggest the following:

Insight 3 (Varying Delivery Structure)**.** *Service providers have varying methods for account reuse in efforts to to distribute suspicion across their account pools.*

Freemium providers have high, 0.8 overlap which results in an estimated pool slightly larger than either set of honeypot followers. However, premium providers have an interesting split which reveals that fastfollowerz and intertwitter have near 1.0 overlap, resulting in pool size roughly equal to each follower set. This indicates near-perfect account reuse across customers – these providers may have small pools, or alternate between sub-pools. Conversely, devumi and twitterboost have near 0 overlap, suggesting that they may each have a single, large fixed pool of usable accounts. By MSE, we estimate that the pool size contains over 55K fake accounts.

b) Inter-Network Patterns: We study the pairwise *inter-network overlap* of followers across providers to analyze if providers share followers.

Table V shows a matrix with the pairwise overlap coefficients, which suggests the following surprising insight:

Insight 4 (Collusion)**.** *Service providers seem to collaborate with and draw from each other to commit fraudulent actions.*

There is substantial overlap within freemium and premium providers. While fastfollowerz and intertwitter share no accounts with the other premium providers, devumi and twitterboost have a .07 overlap. All 4 freemium providers have a large 0.6-0.7 overlap, indicating that most of their users are *the same*. Furthermore, freemium and premium followers do not overlap, evidencing that freemium followers are otherwise real accounts whereas premium followers are synthetic. Nonzero overlap between providers indicates either a willingness to share follower accounts between fraud providers, or commonality in leaked or hijacked accounts. Upon further inspection, we notice a number of other similarities: Overlapping providers shared domain WHOIS protectors, overlapping premium providers used the same SEO plugins and stylesheets, all freemium providers have two-column sites advertising up to 30K followers and priced from £9.99, and all freemium providers contained the footnote *"[service] is Not Affiliated With OR Endorsed By Twitter.com."*.

C. Attribute Observations

In this section, we study the account attributes of fake followers. Table VI shows per-service entropy in bits for sev-

TABLE V: Fraud providers share follower accounts.

		fastfollowerz	intertwitter	devumi	twitterboost	plusfollower	newfollow	hitfollow	bigfolo
Premium	fastfollowerz	1.0	0	0	0	0	0	0	0
	intertwitter	0	1.0	0	0	0	0	0	0
	devumi	0	0	1.0	.07	0	0	0	0
	twitterboost	0	0	.07	1.0	0	0	0	0
Freemium	plusfollower	0	0	0	0	1.0	.65	.69	.64
	newfollow	0	0	0	0	.65	1.0	.64	.63
	hitfollow	0	0	0	0	.69	.64	1.0	.63
	bigfolo	0	0	0	0	.64	.64	.63	1.0

eral follower attributes. These attributes have varying outcome spaces – creation date can be from 2006-2016, booleans have 2 outcomes, and there were 35 language identifiers and 39 UTC settings. For counts, we log-binned the space into 32 bins from 1 to 1M to capture skewed activity levels. Per service, we aggregate attribute values and compute entropy over outcomes. The table shows the empirical sample entropy in addition to the maximum possible entropy. Lower entropy indicates high synchronicity between followers.

The most striking insight from Table VI is as follows:

Insight 5 (Entropy Gap)**.** *Premium service providers deliver followers with low entropy, high regularity attributes, whereas freemium service providers have more attribute disparity.*

Premium providers have much lower entropy in many attributes versus freemium ones, and near 0 entropy in other attributes like geolocation. We elaborate next.

1) Account Creation: devumi, twitterboost and fastfollowerz have very low creation year entropy compared to freemium providers. While both freemium and premium accounts appear to be created more recently (perhaps due to higher suspension rate in older accounts), premium providers have a heavy bias towards recently created accounts (>2014).

2) Profile Defaults: fastfollowerz has a much lower default profile entropy than other providers – we found that >84% of these accounts *did not* have a default profile, whereas default profiles are actually *more common* than not in freemium accounts. Surprisingly, fastfollowerz, devumi and twitterboost also have near 0 entropy for profile image compared to the much higher entropy for freemium providers. We find that premium followers almost always set a custom image, suggesting that the information was fabricated or stolen from real users. Conversely, default profile images are not uncommon for freemium service accounts – this is intuitive, most real users do not fully customize their profiles.

3) Action Counts: devumi and twitterboost have much lower entropy for action counts compared to freemium providers. fastfollowerz also exhibits lower entropy. As Figure 1d shows, there is even more variation between premium providers. Figure 1d shows that intertwitter (P1 "smart") follower counts are disparate and closer to genuine users' entropy (probably due to fraud providers' "camouflage" attempts), unlike other premium fraudsters (P2 "naïve") who behave robotically. Freemium followers have lower follower count entropy compared to genuine ones (and are also more similar to them than premium followers), which is intuitive as while the freemium followers are real, their follower counts are more similar due to the follows traded between themselves. We noticed similar patterns status and favorite counts as well. The lower entropy of action counts in premium providers stems from the variety of packages they sell for Twitter engagement – in addition to fake followers, the premium providers also offer fake retweets and favorites services. Thus, premium providers are incentivized to reuse accounts for multiple types of fraud, and when done naïvely result in high synchrony in "serviceable" attributes.

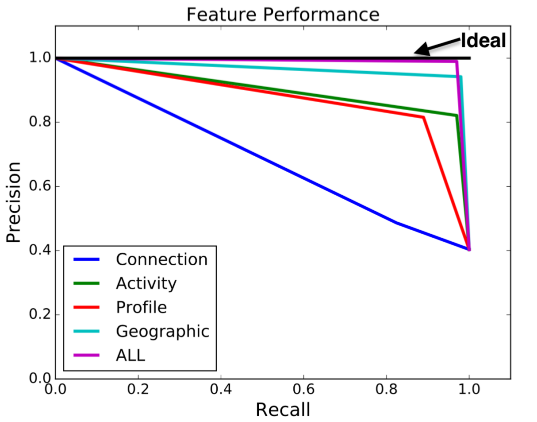


Fig. 4: **Leveraging all features together gives the best detection performance.**

4) User Settings: fastfollowerz, devumi and twitterboost all have near 0 geolocation, language, and tweet protection entropy. Of these, devumi and twitterboost accounts all have US English language, disabled geolocation and unprotected tweets. fastfollowerz has a slightly higher language entropy of .06, but uses exclusively US and GB English, suggesting a heavy premium bias for "English-speaking" accounts. We also find that premium followers mostly have USA timezones. "Smart" intertwitter followers' high language entropy from Figure 1d suggests an aim to better camouflage user attributes compard to the "naïve" providers. Given that intertwitter also has some verified accounts, we hypothesize that the accounts may be hijacked. This is in contrast with freemium followers, which have frequently enabled geolocation, varying languages and protected tweets. Figure 1d also shows that freemium followers appear similar to genuine ones as they are otherwise real accounts. However, we find that freemium followers have higher language entropy than genuine ones, as they are spread over many countries whereas genuine followers tend to disproportionately speak their followee's language.

IV. Discriminative Power of Entropy Features

Can we leverage these differences to discriminate user behaviors? In this section, we evaluate several attribute features

TABLE VI: Per-service entropy (in bits) over account attribute distributions.

	Service	Created (year)	Def. Prof.	Def. Prof. Image	# Favorites	# Followers	# Friends	# Lists	# Statuses	Geolocation	Lang.	Protected	UTC	Verified
Premium	fastfollowerz	1.37	.63	.01	3.65	2.73	2.73	2.99	3.8	.00	.06	.00	1.04	.00
	intertwitter	2.99	.82	.94	4.04	3.54	2.63	2.53	4.31	.67	2.55	.56	1.97	.18
	devumi	1.13	.97	.02	1.05	1.54	1.17	2.49	1.18	.00	.00	.00	1.42	.00
	twitterboost	1.13	.97	.03	1.05	1.56	1.16	2.51	1.15	.00	.00	.00	1.41	.00
Freemium	plusfollower	1.82	.93	.73	4.18	3.76	3.38	2.73	4.40	.54	2.04	.30	1.70	.00
	newfollow	1.68	.90	.75	4.20	3.70	3.32	2.64	4.37	.55	1.99	.28	1.62	.00
	hitfollow	1.78	.93	.73	4.14	3.76	3.32	2.72	4.37	.52	2.01	.30	1.70	.00
	bigfolo	1.88	.92	.75	4.20	3.74	3.34	2.72	4.40	.56	2.05	.32	1.71	.00
	Max Entropy:	3.46	1.00	1.00	5.00	5.00	5.00	5.00	5.00	1.00	5.13	1.00	5.29	1.00

on their supervised discriminative power.

We classified the engineered entropy features from Table VI into the following groups based on feature type: *Connection* (#followers, #friends), *Activity* (#statuses, #lists, #favorites), *Profile* (default profile and image, verified, created at), *Geography* (language, UTC) and *All* (the union of all groups). Note that while we nominally refer to these features as above, they refer to the *entropy of the feature over account followers*, rather than raw values of the account itself.

We evaluate these features using binary classification (genuine vs. fraudulent) as is traditionally done in practice. We use a Support Vector Machine (SVM) with radial basis function (RBF) kernel and 10-fold cross validation. Our ground-truth dataset consists of 307 fraudsters and 200 genuine users. The fraudulent accounts are a combination of premium and freemium honeypots and accounts whose profiles have been listed on freemium providers' websites as service users. Genuine accounts belong to well-known academics in data mining.

Figure 4 shows the relative performance of our feature groups in terms of overall precision and recall. We notice that *Connection* features perform comparatively poorly, *Profile* and *Activity* features perform better, *Geography* performs even better, and the combination *All* performs near-ideal with .98 precision even with .95 recall (better recall than supervised approaches which use raw account features [19]), indicating strong performance.

V. Conclusion

In this work, we aimed to study the nature of link fraud. To this end, we setup honeypot accounts, purchased fake followers for them from several fraud services, and carefully instrumented a data scraping process to capture their behaviors. Specifically, we made the following contributions:

- **Observations**: After carefully collecting and analyzing ground-truth link fraud data, we present numerous new insights about different types of fraudsters, their mission constraints and behaviors.
- **Features**: From these insights, we engineer novel follower-centric entropy features that allow us to accurately differentiate between fraudsters and genuine users (>.95 precision and recall).

VI. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. CNS-1314632, IIS-1408924, and by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties.

References

[1] A. Beutel, W. Xu, V. Guruswami, C. Palow, and C. Faloutsos, "Copycatch: stopping group attacks by spotting lockstep behavior in social networks," in *WWW*, 2013.

[2] B. A. Prakash, A. Sridharan, M. Seshadri, S. Machiraju, and C. Faloutsos, "Eigenspokes: Surprising patterns and scalable community chipping in large graphs," in *PAKDD*. Springer, 2010, pp. 435–448.

[3] N. Shah, A. Beutel, B. Gallagher, and C. Faloutsos, "Spotting suspicious link behavior with fbox: An adversarial perspective," in *ICDM*, 2014.

[4] D. M. Freeman, "Using naive bayes to detect spammy names in social networks," in *AISec*. ACM, 2013, pp. 3–12.

[5] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@ spam: the underground on 140 characters or less," in *CCS*. ACM, 2010, pp. 27–37.

[6] 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*, 2013.

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

[8] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, "Detecting and characterizing social spam campaigns," in *SIGCOMM*, 2010.

[9] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy," in *WWW*, 2013.

[10] N. Perloth, "Fake twitter followers become multimillion-dollar business," April 2013, [Online; posted 5-April-2013]. [Online]. Available: http://bits.blogs.nytimes.com/2013/04/05/fake-twitter-followers-becomes-multimillion-dollar-business/

[11] 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.

[12] G. Stringhini, G. Wang, M. Egele, C. Kruegel, G. Vigna, H. Zheng, and Y. B. Zhao, "Follow the green: Growth and dynamics in twitter follower markets," in *SIGMETRICS*, 2013.

[13] A. Aggarwal and P. Kumarguru, "What they do in shadows: Twitter underground follower market," in *PST*, 2015.

[14] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on twitter," in *CEAS*, 2010.

[15] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots + machine learning," in *SIGIR*. ACM, 2010, pp. 435–442.

[16] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *ACSAC*. ACM, 2010, pp. 1–9.

[17] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *CCS*. ACM, 2014, pp. 477–488.

[18] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Catchsync: catching synchronized behavior in large directed graphs," in *KDD*. ACM, 2014, pp. 941–950.

[19] M. Mccord and M. Chuah, "Spam detection on twitter using traditional classifiers," in *ICATC*. Springer, 2011, pp. 175–186.