Revealing Fake Profiles in Social Networks by Longitudinal Data Analysis

Aleksei Romanov, Alexander Semenov and Jari Veijalainen
University of Jyväskylä, Finland

Keywords: Social Network Analysis, Anomaly Detection, Suspicious Behaviour, Graph Mining, Longitudinal Data.

Abstract: The goal of the current research is to detect fake identities among newly registered users of vk.com. Ego networks in vk.com for about 200.000 most recently registered profiles were gathered and analyzed longitudinally. The reason is that a certain percentage of new user accounts are faked, and the faked accounts and normal accounts have different behavioural patterns. Thus, the former can be detected already in a few first days. Social graph metrics were calculated and analysis was performed that allowed to reveal outlying suspicious profiles, some of which turned out to be legitimate celebrities, but some were fake profiles involved in social media marketing and other malicious activities, as participation in friend farms.

1 INTRODUCTION

Social media sites started to appear around 2005 and many of them have attracted hundred of millions of users. The number of distinct profiles at Facebook exceed one billion. Because social media sites want to attract as many users as possible, strong authentication of user's identity is not required by them when a new user joins the site. The sites usually require in their EULA that real persons, associations and companies must use their true identity in their profile. Some sites, like Twitter, also allow so-called parody accounts or profiles, where parts of a real user's identity, such as name, image, email address, etc., can be utilized but the profile must clearly state in the description that it is a parody profile. For the authentication at many sites it is usually enough that a user has a browser, internet connection, and a functioning email address and/or functioning phone number that can be used to send a verification link or code back from the site. It must then fed into the browser while finalizing the profile creation. The service provider has IP-addresses that were used while the account of profile was established, but these can be dynamically allocated, or refer to computers in a shared use. Thus, through them the identity of the real profile owner cannot be established. Further, email accounts can be easily established at service providers, such as gmail, hotmail etc., again without strong authentication, and prepaid SIM-cards obtained without identification. Thus, there are numerous profiles and account at various sites that are in some sense misleading or false. These include stolen identities of existing people (duplicates) that might or might not have a profile at the site in question, but also fake identities that are, for instance, combining a picture of a real person to other, fabricated credentials. A further case are compromised accounts or profiles where the original user has lost control over the profile or account and it is used by perpetrators for various, mostly criminal, or in any case questionable purposes. Facebook annual report says, that 5,5% – 11,2% of worldwide monthly active users (MAUs) in 2013-2014 were false (duplicate, undesirable, etc.) (Facebook, 2014). Because the perpetrators can hide their true identity, false identities (also referred to as "sybils") play an important role at initial phases of advanced persisted threats (APT), phishing, scam, or other forms of fraud and malicious activities. One of the recent trends is crowdturfing - the term representing a merger of astroturfing (i.e., sponsored information dissemination campaigns obfuscated to appear spontaneous movements) and crowdsourcing. For instance, according to the study by Harvard Business School, popular site Yelp.com filters 16% of reviews as fake; in the end of 2015 Amazon.com has started legal action against more than 1.000 unidentified people it claims provide fake reviews through Fiverr platform on the US version of its website (Gani, 2015).

The largest European online social media site, which is especially popular in Russia and in post-

Romanov, A., Semenov, A. and Veijalainen, J.
Revealing Fake Profiles in Social Networks by Longitudinal Data Analysis.
DOI: 10.5220/0006243900510058
In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST 2017)*, pages 51-58
ISBN: 978-989-758-246-2

Soviet countries is vk.com: in October 2016 it had around 390 million registered users, and it was ranked 14th in global Alexa.com web-sites ranking. It has its servers in Russia.

Each user of vk.com has unique numeric identifier. These identifiers have been allocated in an (almost strict) ascending order with the advancing registration time. Therefore, it becomes possible to assort the profiles, approximately, on a timeline, taking into accounts identifiers' ordering. One can access any account by using its identifier; if the account does not exist, vk.com would return an error message. Thus, it is possible to find the most recently registered accounts and follow their activity using data collection software. For instance, it is possible to follow the development of the friend network and contents propagated by recently established accounts over time that allows interesting temporal data analysis.

There are internal security mechanisms in vk.com that freeze or deactivate profiles that get a number of reports for abuse, spam or fraud activities. There isn't much information available on this topic. The system mainly relies on other users' amount of reports and then automated or manual analysis by administrators. The fake and malicious profiles are deactivated with time, but the problem is that it's unlikely that they will be defined as fake or malicious unless they start their attacks, and some time is also needed for administrators to react on the reports. The time gap between attack's start and deactivating the profile can be enough for the fraudster to achieve the aim of attack. What we are interested in is to detect fake accounts before they initiate the main phase of the attack on the stage of preparation. The information about banned or deleted state of an account through time can be treated as a ground truth that the profile was indeed fake or malicious.

Our research aims at detecting fake accounts at online social media sites using longitudinal data analysis. Because of the features described above, we have chosen vk.com to become our target. In particular, the goal of the current research is to detect fake identities among newly registered users of vk.com. We hypothesize, that a certain percentage of new user accounts are fake. The fake accounts and normal accounts have different patterns in these respects and thus the former can be detected already in a few days.

The **aim** of this paper is in providing descriptive analysis of longitudinal data collection process of 200.000 newly registered users of vk.com and testing the following **hypothesis**: fake profiles are more likely to be found among those users that show abnormal behaviour in growth of social graph metrics such as degree, reciprocated ties and clustering.

The paper is structured as follows: section 2 describes related work, section 3 details notation and collected data, section 4 explains the social graph metrics that were considered, section 5 provides analysis results, and section 6 concludes the paper.

2 RELATED WORK

There is a number of research papers aimed at detecting false identities in social media, and their manifestations such as fake reviews on review sites, and spam reviews. Majority of methods are based on extraction of various features from profiles and messages, and then machine learning algorithms are used in order to build a classifier capable of detecting false accounts based on extracted features. Some work was done on developing algorithms for detecting simultaneous liking of particular pages on Facebook by a group of fake profiles or paid users.

The authors of (Beutel et al., 2013) use graph based approach to detect attackers with lockstep behaviour – users acting together in groups, generally liking the same pages at around the same time. The algorithm called CopyCatch that operates similarly to mean-shift clustering with a flat kernel (Cheng, 1995) is actively used at Facebook, searching for attacks on Facebook's social graph that enables to limit "greedy attacks". The authors face the problem of co-clustering pages (subspace clustering) and likes (density-seeking clustering) at the same time, which is known as a NP-hard problem and is often solved by approximation techniques. That is why two algorithms are presented – one provably-convergent iterative algorithm and one approximate, scalable MapReduce (Dean and Ghemawat, 2008) implementation.

In the article (Ikram et al., 2015) fraudulently boosting the number of Facebook page likes using like farms was addressed. In contrast to the CopyCatch algorithm mentioned above the authors incorporate additional profile information to train machine learning classifiers. They characterized content generated by social network accounts on their timelines, as an indicator of genuine versus fake social activity. They then extracted lexical and non-lexical features and showed that like farm accounts tend to often re-share content, use fewer words and poorer vocabulary, and more often generate duplicate comments and likes compared to normal users. Further, a classifier was built that allowed to detect known like farm accounts (boostlikes.com, authenticlikes.com, etc.) with high accuracy.

It is known that fraudsters may be paid to disguise certain account to seem more trustworthy or popu-

lar by artificial involvement of additional followers. Such service is supplied by fake accounts or through real accounts hijacked with malware. This phenomenon creates distorted images of popularity and legitimacy, with unpleasant or even dangerous effects to real users. The authors of the recent paper (Jiang et al., 2016) focus on synchronised behaviour detection and present CATCHSYNC algorithm.

One other article that touches a question of revealing camouflaged behaviour is (Hooi et al., 2016) mainly focusing on a Twitter dataset.

Adicari (Adikari and Dutta, 2014) describes identification of fake profiles in LinkedIn. The paper shows that fake profile can be detected with 84% accuracy and 2.44% false negative, using limited profile data as input. Such methods as neural networks, support vector machines, and principal component analysis are applied. Among others, such features as number of languages spoken, education, skills, recommendations, interests, and awards are used. Characteristics of profiles, known to be fake, and posted on special web sites are used as a ground truth.

The paper by Chu et al. (Chu et al., 2010) aim at differentiating Twitter accounts operated by human, bots, or cyborgs (i.e., bots and humans working in concert). As a part of the detection problem formulation, the detection of spamming accounts is realized with the help of an Orthogonal Sparse Bigram (OSB) text classifier that users pairs of words as features. Accompanied with other detecting components assessing the regularity of twits and some account properties such as the frequency and types of URLs and the use of APIs, the system was able to accurately distinguish the bots and the human-operated accounts.

In addition to, or instead of analyzing the individual profiles, another stream of approaches rely on graph-based features when distinguishing the fake and legitimate accounts. For instance, in the paper (Stringhini et al., 2010) methods for spam detection in Facebook and Twitter are described. The authors created 900 honeypot profiles in social networks, and performed continuous collection of incoming messages and friend requests for 12 months. User data of those, who performed these requests were collected and analyzed, after which about 16.000 spam accounts were detected. Authors further investigated the application of machine learning for further detection of spamming profiles. On top of the features used in the studies above, the authors were also using the message similarity, the presence of patterns behind the search of friends to add, and the ratio of friend requests, and then used Random Forest as a classifier.

A paper (Conti et al., 2012) proposes an application of graph features for the detection of fake profiles. The authors base their detection method on analysis of distribution of number of friends over time. Boshmaf et al. (Boshmaf et al., 2016), however, claim that the hypothesis that fake accounts mostly befriend other fake accounts does not hold, and propose new detection method, which is based on analysis features of victim accounts, i.e. those accounts, which were befriended by a fake account.

The structure of the social graph of active Facebook users and numerous features were studied in the paper (Ugander et al., 2011). However, the research was done only for one data snapshot.

Like farms were studied thoroughly, however little studies were targeted on revealing friend farms. Moreover, there are very few research papers that analyzed the behaviour of users longitudinally, crawling data periodically and analyzing them in order to capture the dynamics. In this research we are doing this.

3 GATHERED DATA

3.1 Notation

We use the following notation: graph of a social network $G = (V,E)$ consists of a set of vertices $V = \{v_1,...,v_n\}$ and a set of m edges $E \subset V \times V$, $|V| = n$ and $|E| = m$. If $(v_i,v_j) \in E$, then vertices v_i, and v_j are called adjacent. If every of two vertices are adjacent, the graph is called complete. Neighbourhood $\mathcal{N}(v)$ of a vertex v is a set of all vertices v' adjacent to v, i.e. $v' \in \mathcal{N}(v)$ for all $(v,v') \in E$. Then, the *degree* of v, $deg(V) = |\mathcal{N}(v)|$. Path $\mathcal{P}_{ij}$ between vertices i and j is a sequence of vertices $v_0,...,v_d$ such that $v_0 = v_i$, $v_d = v_j$, and $(v_k,v_{k+1}) \in E$, $\forall k = 0,...,d-1$. Such path is called a path of length $d-1$. Two vertices v_i and v_j are connected, if there is a path between them. Graph G is connected, if all of its vertices are pairwise connected, and disconnected otherwise.

A graph can be represented with an adjacency matrix, which is a matrix with rows and columns labeled by graph vertices, with a 1 or 0 in position (v_i,v_j) according to vertices' v_i and v_j adjacency property. A graph without self-loops has zeros on the diagonal. For an undirected graph, the adjacency matrix is symmetric. In our case, we have no self-loops undirected graph. For example, adjacency matrix for graph on figure 1 can be found in equation 1.

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad (1)$$

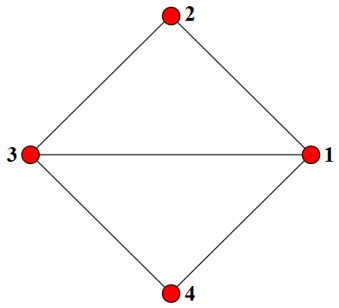


Figure 1: An example of an adjacency matrix.

For any subset of vertices $S \subseteq V$, $G[S] = (S, (S \times S) \cap E)$ denotes the *subgraph*, a group, induced by S on G. A vertex belonging to S is referred to as a group vertex, vertices in $V \setminus S$ are considered to be the non-group vertices. The group $G[S]$ is called a *clique* if the *subgraph* induced by S is complete.

In this context the user profiles hosted by vk.com are modelled by vertices of the graph, and "friendship" relations are modelled by undirected edges between the vertices. Two user profiles are in this relationship, if they both $follow$ each other, according to the site ontology of vk.com.

Given a graph $G = (V(G), E(G))$ an *induced subgraph* of G, $G_s = (V(G_s), E(G_s))$, is a graph that satisfies the following conditions:

$$V(G_s) \subset V(G), E(G_s) \subset E(G),$$
$$\forall u, v \in V(G_s), (u,v) \in E(G_s) \Leftrightarrow (u,v) \in E(G).$$

When G_s is a induced subgraph of G, it is denoted as $G_s \in G$.

The *neighbourhood subgraph* of radius r of vertex v is the subgraph induced by the neighbourhood of radius r of v and denoted as $G_s^r(v)$.

An *ego network* is a neighbourhood subgraph of radius 1 of vertex v, $G_s^1(v)$ or just $G_e(v)$. In other words, such subgraph that consists of one "focal" vertex, the vertices to which ego is directly connected, and edges between these vertices. More information on ego networks can be found in the paper (Freeman, 1982).

The attributes that are included into the vertices of the graph are: a) id – unique identifier that was generated by vk.com and that each user obtains after registration; b) first and last name; c) gender; d) city which is represented as city_id and the real city name is acquired through API request; e) status which a user can post right below his or her name and if the user is fake or malicious that's the first place where a link to malware is (usually) put f) timestamp of the last activity by which we can understand whether the profile is registered and abandoned or active every day, its temporal activity.

A timestamp denotes a date and a time when the exact data gathering was made, i.e. t_1 ="2016-05-12 08:04:33" or t_2 ="2016-05-12 10:49:18".

$$\mathbb{T} = [t_1, t_{2,\ldots}, t_N]$$

The time interval between two timestamps is $\sim$2 hours and number of timestamps is $N \in [1, 55]$, thus we cover around 5 days.

For each timestamp an ego network is gathered for every user u from 200.000 of the targeted group. A user is represented in the ego network as a focal vertex v_u. Thereby, we obtain the following sequence of ego networks evolving through time:

$$G_e(v_u) = \{G_{et_1}(v_u), G_{et_2}(v_u), \ldots, G_{et_N}(v_u)\}$$

Further, social graph metrics are calculated and analysed for every ego network (section 4 and 5).

3.2 DATA COLLECTION

We have developed a crawler capable of performing longitudinal collection of ego networks of set of vk.com users $V = \{v_1, \ldots, v_n\}$. Then, we have identified the most recently registered profiles and performed longitudinal crawling of their ego networks and user details. The first data collection was gathering ego networks of 200.000 newly registered users for each 2 hours during the period of 5 days. The collection was performed in May 2016, the next one was functioning in a similar manner, but the collection lasted for nearly 1 month, during September 2016. The latter collection consisted of over 5 TB of data on the disk.

Data: set of vk.com users $V = \{v_1, \ldots, v_n\}$
Result: set $\{G_e(v_1), \ldots, G_e(v_n)\}$
initialisation;
forall $user_i \in V$ **do**
  collect friends F_0 of $user_i$;
  forall $user_j \in F_0$ **do**
    insert edge (v_i, v_j) into $G_e(v_i)$;
    collect friends F_1 of $user_j$;
    forall $user_k \in F_1$ **do**
      if k *in* F_0 **then**
        insert edge (v_j, v_k) into $G_e(v_i)$;
      end
    end
  end
end

Algorithm 1: Longitudinal crawling.

We filtered ∼11.000 users that started their activity from 1st snapshot to 2nd and then calculated social graph (graph of the social network) parameters for the ego networks they form with their friends and friends of friends. Activity in this case means registration and launch of adding friends (i.e. user in 1st snapshot had 0 friends or tagged as not created, but in the 2nd snapshot – created and with $n > 0$ friends).

4 SOCIAL GRAPH METRICS

There are different metrics of social graphs (i.e. centrality, degree, closeness, etc.). We focus our attention on the following features of the social graph formed based on the collected dataset.

4.1 Degree Distribution

We understand degree as the number of reciprocated ties (friendship) for each node (user).

Growth of vertices' degree: we have found, that the degree of a number of vertices grows very quickly; majority of the user profiles modeled by these vertices belong to celebrities.

4.2 Reciprocated Ties

The number of transitive triplets $\sum_{i,h} x_{ih}x_{ij}x_{jh}$, where x are elements of adjacency matrix A (2) corresponding to the graph and i is fixed to the current focal vertex of the ego network.

$$A = \begin{pmatrix} x_{00} & x_{01} & \cdots & x_{0n} \\ x_{10} & x_{11} & \cdots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{nn} \end{pmatrix} \quad (2)$$

4.3 Clustering

Clustering is calculated as follows:

$$c_{v_u} = \frac{2T(v_u)}{deg(v_u)(deg(v_u)-1)}, \quad (3)$$

where $T(v_u)$ is the number of triangles through vertex v_u and $deg(v_u)$ is the degree of v_u.

5 ANALYSIS

We have found, that some of the accounts that demonstrate unusually high clustering coefficient, same time having large number of friends (e.g. nearly 150 friends, and $c_u = 1$, meaning that ego network forms a clique, i.e. all of the nodes are connected) use "friend farm" services that allow them to gain large number of friends in a short time.

Firstly, we take a look at overall degree distribution for 187.803 active users in the timestamp 55, which is presented on figure 3. The weighted mean $\bar{x} = 5,43$, standard weighted deviation $sd_w = 22,48$. The red line denotes $3sd_w = 67,44$, there are 275 users that have more than 67 friends and are treated as suspicious.

Then we narrow down our sample, and figure 4 represents clustering distribution for 2.846 users in timestamp 55 that have more then 3 friends. 3% of users have clustering $> 0,8$. The average clustering $\bar{c} = 0,25$.

We go back to 11.000 filtered users who started their activity between the 1st and the 2nd snapshots. Figure 5 shows the cumulative degree distribution among these users. 1.760 people who have added friends have 2,8 friends in average in the first snapshot. 4.500 users with at least one friend have 7 friends in average by snapshot 10. Firstly, the speed of adding friends was high, but it slows down by snapshot 10. The average speed of adding friends is shown in table 1.

Table 1: Change in average speed of adding friends from snapshot 2 to 10.

Snapshot		Average speed
From	To	of adding friends
2	3	2,86
3	4	1,20
4	5	0,86
9	10	0,10

Clustering for the filtered users is presented on figure 6. There is a peak 0,5 – 0,6 range and one more in 0,8 – 0,9 range. We are interested in users with clustering higher than 0,8, because they tend to form cliques.

Figure 2 shows relationship between friends and clustering. There could be found profiles with unusually high number of friends and high clustering, which is considered to be suspicious and such users are more likely to be involved in "friend farms". A group where a user can post a message that he or she is inviting other users to establish artificial "friendship" relations. The main idea behind a *friend farm* is to gain a number of random "friends", which are not actually friends. Many of users in such communities are usually bots or fake accounts.

Some accounts have a lot of friends, and very low clustering, that means that their friends do not "know"

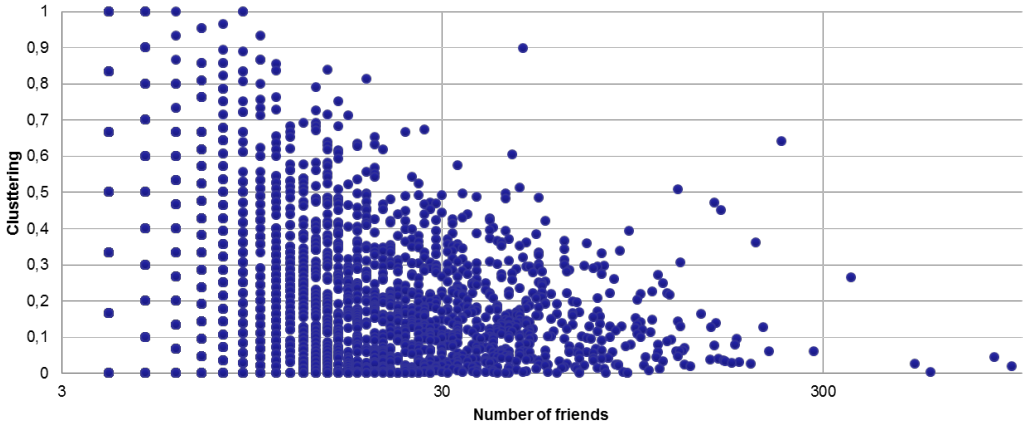


Figure 2: Clustering and friends ratio, approx. 3 days after registration.

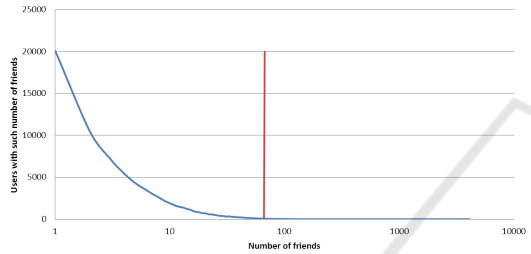


Figure 3: Degree distribution for 187803 registered users in timestamp 55.

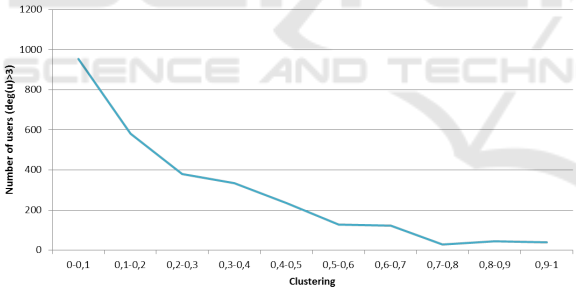


Figure 4: Clustering distribution for timestamp 55 for all users with degree > 3.

Table 2: Growth in number of friends and number of triangles for a user with high clustering (*TS – timestamp).

TS*	# of Friends	# of Triangles	Clustering
3	3	3	1,00
4	17	134	0,99
10	44	864	0,91
30	47	999	0,92
55	49	1055	0,90

each other, so perhaps they add random people.

Table 2 represents one of the real life evolution of a user with id 364712485. A set of figures 7 visualize his ego network respectively to the table. We do not provide statistics for each timestamp for practical reason to save space, and visualisation becomes unsuitable for more number of friends and triangles.

We considered the chosen user as a suspicious one and analyzed the content of his web page. This exact user was a member of several friend farms (fi-

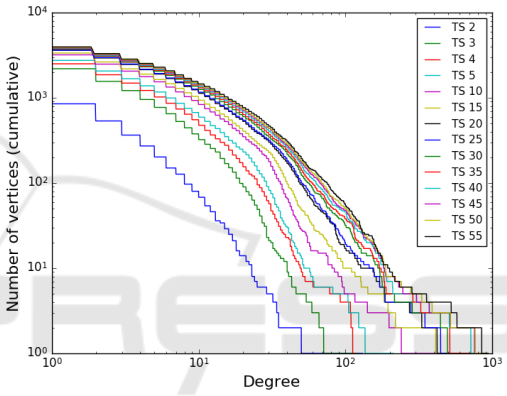


Figure 5: Degree distribution for 11.000 filtered users in timestamps 2 to 55.

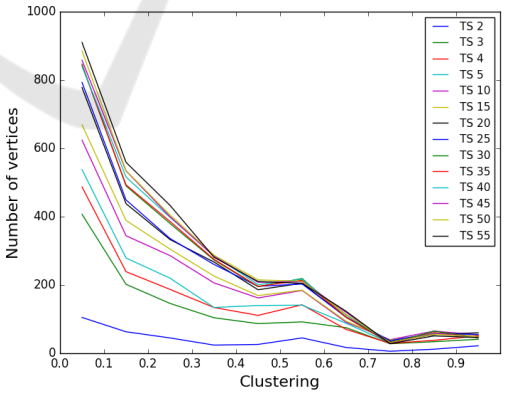
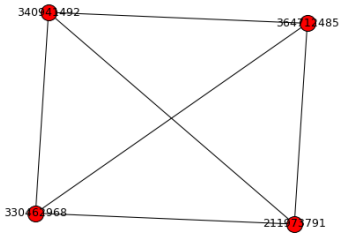
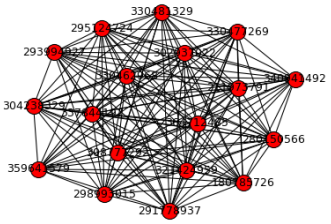


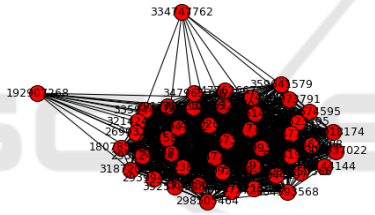
Figure 6: Clustering for 11.000 filtered users in timestamps 2 to 55.



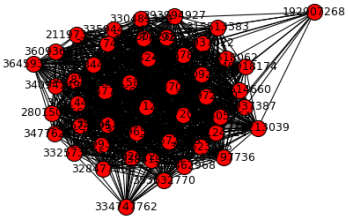
(a) Timestamp 3.



(b) Timestamp 4.

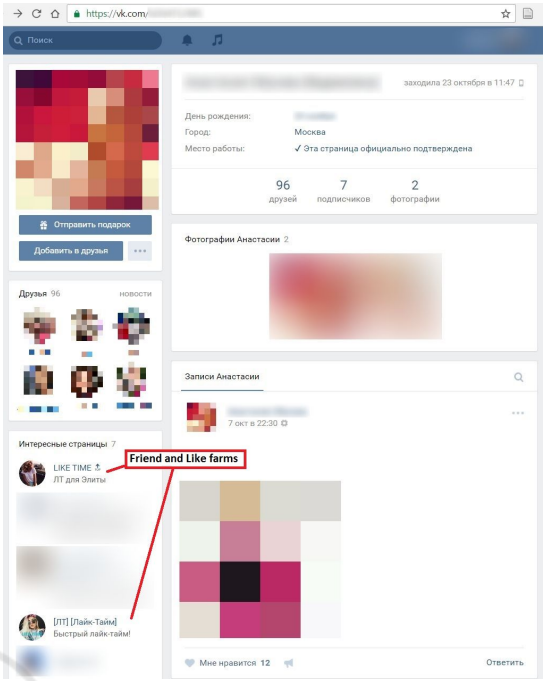


(c) Timestamp 10.

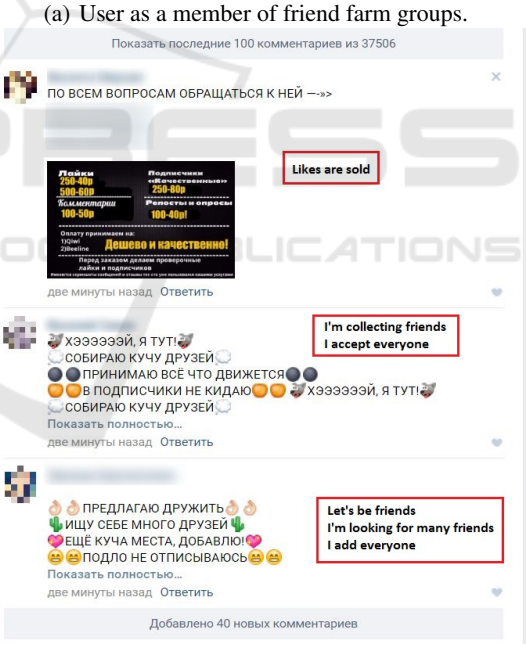


(d) Timestamp 30.

Figure 7: A real life evolution example of highly clustered ego network.



(a) User as a member of friend farm groups.



(b) Comments section of a friend farm group.

Figure 8: Content analysis of suspicious user and group.

gure 8(a)). That was the time when we discovered friend farms. An example of a comments section in one of such group can be found on figure 8(b), where users claim to add anyone to friends who will send them a request.

6 CONCLUSION AND FURTHER RESEARCH

A longitudinal collection of ego networks in vk.com was done for about 200.000 the most recently registered profiles for each 2 hours during the period of 5

days. One more collection was performed later that lasted for nearly 1 month and occupies 5 TB of disk space for the further research. We took a look at overall state of the gathered social graph by calculating weighted mean, standard weighted deviation and found suspicious outlying users. Than, for the filtered ∼11.000 users that started their activity from 1st snapshot to 2nd we calculated social graph parameters (degree, reciprocated ties and clustering) for the ego networks that they form with their friends and friends of friends. The analysis of suspicious users allowed us to reveal *fake profiles* and even *friend farms*, that we are going to study in more details in future research. Hence, we **accept the stated hypothesis** that fake profiles are more likely to be found among those users that show abnormal behaviour in growth of social graph metrics. The **contribution** of this paper is in the descriptive analysis of vk.com users' longitudinal data collection, accepting the stated earlier hypothesis and revealing "friend farms".

The further research is aimed on consecutive immersion in friend farm groups. We will focus on users who are active in this groups and analyze their actions through time to understand their behavioural strategy of gaining new friends. Then we would be able to answer the question whether the friend farms are an efficient instrument or not to make a profile look less suspicious for subsequent implementation of advanced persistent threats.

We have also identified number of websites which sell fake social media accounts (including vk.com and other sites). One of the further research directions is to purchase several accounts as a ground truth about fake profiles and analyze their behaviour (highly likely, they would be in our dataset already), compare their characteristics with legitimate accounts.

REFERENCES

Adikari, S. and Dutta, K. (2014). IDENTIFYING FAKE PROFILES IN LINKEDIN. *PACIS 2014 Proceedings*.

Beutel, A., Xu, W., Guruswami, V., Palow, C., and Faloutsos, C. (2013). CopyCatch: Stopping Group Attacks by Spotting Lockstep Behavior in Social Networks. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 119–130, New York, NY, USA. ACM.

Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M., Beznosov, K., and Halawa, H. (2016). Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs. *Computers & Security*, 61:142–168.

Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790–799.

Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2010). Who is Tweeting on Twitter: Human, Bot, or Cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 21–30, New York, NY, USA. ACM.

Conti, M., Poovendran, R., and Secchiero, M. (2012). FakeBook: Detecting Fake Profiles in On-Line Social Networks. In *ResearchGate*, pages 1071–1078.

Dean, J. and Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM*, 51(1):107–113.

Facebook, i. (2014). Facebook annual report, FB-12.31.2014-10k.

Freeman, L. C. (1982). Centered graphs and the structure of ego networks. *Mathematical Social Sciences*, 3(3):291–304.

Gani, A. (2015). Amazon sues 1,000 'fake reviewers'. *The Guardian*.

Hooi, B., Song, H. A., Beutel, A., Shah, N., Shin, K., and Faloutsos, C. (2016). FRAUDAR: Bounding Graph Fraud in the Face of Camouflage. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 895–904, New York, NY, USA. ACM.

Ikram, M., Onwuzurike, L., Farooqi, S., De Cristofaro, E., Friedman, A., Jourjon, G., Kaafar, M. A., and Shafiq, M. Z. (2015). Combating Fraud in Online Social Networks: Detecting Stealthy Facebook Like Farms. *arXiv:1506.00506 [cs]*. arXiv: 1506.00506.

Jiang, M., Cui, P., Beutel, A., Faloutsos, C., and Yang, S. (2016). Catching Synchronized Behaviors in Large Networks: A Graph Mining Approach. *ACM Trans. Knowl. Discov. Data*, 10(4):35:1–35:27.

Stringhini, G., Kruegel, C., and Vigna, G. (2010). Detecting Spammers on Social Networks. In *Proceedings of the 26th Annual Computer Security Applications Conference*, ACSAC '10, pages 1–9, New York, NY, USA. ACM.

Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. (2011). The Anatomy of the Facebook Social Graph. *arXiv:1111.4503 [physics]*. arXiv: 1111.4503.