# Twitter Influencers or Cheated Buyers?

Savvas Zinonos, Andreas Tsirtsis and Nicolas Tsapatsoulis
Dept. of Communication and Internet Studies
Cyprus University of Technology,
CY-3036, Limassol, Cyprus
{*sk.zinonos, ai.tsirtsis*}*@edu.cut.ac.cy, nicolas.tsapatsoulis@cut.ac.cy*

***Abstract*—Twitter is one of the most popular social networking platforms that people use to communicate and interact. Organisations and companies use Twitter, as well as other social media platforms, for the marketing of their products or services. To achieve this goal they seek to partner with influential Twitter users, as a part of their influencer marketing strategy. Influencer marketing is considered more effective than traditional marketing. Influencers are more trustworthy than a business due to the fact that they have developed close connection with their followers. This marketing trend has played an important role in the rise of fake influencers in Twitter. Fake influencers inflate their follower counts by buying fake Twitter accounts from vendors and they manage to partner with companies. However, that partnership does not benefit companies as the influencer's engagement is fake. In this paper we analyse centrality and overall network characterization measures applied on Twitter fake influencer accounts and on legitimate influencer accounts. The results showed that the measures we propose are statistically significant and can be easily applied to automatically detect fake influencers on Twitter.**

***Index Terms*—Twitter fake influencers, centrality measures, reciprocity, centralization, influencer marketing, network characterization measures.**

## I. Introduction

The huge expansion of Online Social Networks (OSN) provides a variety of opportunities for communication, marketing and other activities among users, companies and organisations in a global scale. Social media, especially Twitter, gained the trust of consumers by connecting with them at a deeper level [1]. Big companies and brands need to have a strong presence in social media marketing in order to promote their products and services. To do so they need to reach influential users. Twitter influencers are well connected accounts as they follow and are followed by hundreds of other users. They are considered as experts in their area and they can be trusted by other users. Becoming a social media influencer is hard to achieve as it needs a lot of effort and time. With the emergence of vendors selling Twitter fake accounts people can inflate their follower counts, even their engagement. Brands seeking to expand their financial circles through Twitter influencers may partner with fake influencers. Fake influencers secure sponsorships but in reality they cannot impact real users.

Fake influencers are present across all major social networking platforms (Instagram, Twitter, Facebook). Fake influencers have high engagement and community size. Influencer marketing[1] has increased in popularity and the ability to spot fake influencers has become critical to the continued success and credibility of the market. Although some methods were proposed to detect fake Twitter accounts little (if any) effort has been devoted to the detection of fake influencers.

According to the recent literature [2] discovery of fake users in Twitter is usually done using methods that utilise specific account features like date of account creation, number of tweets, number of followers etc, while a fewer number of approaches use centrality measures [3, 4]. Centrality measures have been applied in the past in numerous research works [5, 6] in order to examine influence patterns in inter-organisational networks, to study the power in organisations and analyse the structure of criminal networks.

In this work we emphasise on the detections of fake Influencers in Twitter. For this purpose we evaluate focuses on centrality measures that have not been proposed so far in any other research. Our hypothesis is that both centrality measures and network characterisation measures differ significantly between legitimate and fake influencers. To the best of our knowledge this is the first time such an investigation takes place. The egocentric Twitter networks of both legitimate and fake influencers are freely available to the research community and we hope that this dataset will be utilised by other researchers. The current work would be a first step towards creating a fake influencer score combining the measures that show significant difference between legitimate and fake influencers.

## II. Related Work

Chu *et al.* [7], have studied the problem of automation by bots and cyborgs on Twitter. They collected 500.000 Twitter users through Twitter API[2], with more than 40 million tweets. The authors analyzed the profiles of the Twitter users and classified them into three categories: Human, bot and Cyborg. They classified users by manually checking their user logs and homepages. The user was annotated as a human if they could obtain evidence of original, intelligent, specific and human-like contents. The criteria the authors used for labelling a user as a bot were: the lack of intelligent or original content, the excessive automation of tweeting, the abundant presence of spam or malicious URLs, the aggressive following

[1]https://sproutsocial.com/insights/influencer-marketing/
[2]https://developer.Twitter.com/

behaviour and posting unrelated tweets. They classified users as cyborgs if there was evidence of both human and bot participation. Based on a series of measurements and characterization, researchers trained a Random Forest classifier, which successfully determined the category of a new account.

Zheng *et al.* [8] and [9], focused on automatic detection of fake followers in *Sina Weibo*[3]. The *Sina Weibo* is a social networking site, which could be considered as a Twitter competitor in China. The authors bought $20.000$ fake followers in *Sina Weibo* platform, from four different vendors. The collection of legitimate users, was accomplished with the help of $114$ volunteers. A data-set of $15.000$ accounts was collected. The researchers also crawled $6.472$ celebrity accounts. The authors after a comprehensive analysis of data-set, extracted features from fake and legitimate followers in order to build a SVM classifier. These features are content-based and user-based. As far as content-based features concerned, the authors randomly selected spam and non-spam messages, comments and likes. The user-based features they analyzed included: number of followers for each user, the account creation date and the number of average URLs for each user message.

In a similar study Zhang and Lu [10], investigated, also, the discovery of fake followers in *Sina Weibo*. They developed a sampling-based approach in order to detect fake (accounts) followers. The researchers discovered that fake followers accounts, which are bought by an online vendor have similar followers graph. They made use of a technique in order to extract a portion of the graph and to find the similarities. The researchers performed clustering by dividing the fake accounts followers into $35$ classes and observed the properties of each account.

Cresci *et al.* [4] also focused on detection techniques for fake followers on Twitter. They purchased fake followers from various vendors while the dataset of legitimate followers was collected from *@TheFakeProject*. A series of features were selected and used in order to train the Random Forest Classifier: a) the number of friends, b) the number of tweets, d) the content of tweets and e) the relation between the number of friends and followers. A dataset was created and used to test some suggested approaches. The results of their study showed that the classifier was able to classify correctly the accounts, with high accuracy.

Dickerson *et al.* [11], have based their study on the fact that the collection of tweets sample is limited and the classification of Twitter users as bots or humans is less or not effective with network-based methods. The researchers, presented the *SentiBox* framework, a very interesting sentiment analysis approach. The *SentiBox* has been used to classify Twitter accounts as fake or legitimate. The classification was achieved based on features such as tweets semantics and user behaviour. The results showed that most bots post tweets with positive sentiment, while in the case of human, the number of positive sentiment tweets was much lower.

[3]https://en.wikipedia.org/wiki/Sina_Weibo

Mehrota *et al.* [3] have devised a method to detect all the fake followers within a social graph. This method was based on network features related to the centrality of all the nodes in the graph. The data-set they used in this study was compiled by the study of Gresci *et al.* [4]. The dataset consisted of fake accounts and legitimate accounts. Also, the authors bought fake followers from different vendors. Various centrality measures have been computed, using *networkx*[4], a *Python*[5] library for network analysis. The researchers trained and tested different classifiers: a) Artificial Neural Networks, b) Decision Tree Classifier and c) Random Forest Classifier. The proposed method has shown promising results.

Stringhini *et al.* [12] have focused on the detection of fake followers on Twitter. The authors found some vendors, who sell real accounts as followers to users or to organizations. These vendors, known as *Merchants Pyramid*, in addition to real followers, they also offer fake followers for free within limited time periods. In return, vendors gain access to the user's account, which creates security risks for the account and personal data. Researchers used bought fake followers for the collection of data. About the data set for legitimate users, two million Twitter users were gathered with random data collection. In addition, two more million users were added, with more than $100$ followers. Clustering techniques were used, and an analysis was performed on the dynamics of these fake followers, as well as their detection using user relationship features such as the number of followers and likes, and if these accounts have more followers than friends.

Cao *et al.* [13], introduced a tool, called *SybilRank*, which can detect fake accounts in Social Networks. The researchers, observed that fake profiles, connect to other fake profiles, rather than the legitimate accounts. The tool, relies on social graph properties, in order to rank users according to their perceived likelihood of being fake. The tool was tested on the complete *Tuenti*[6] social graph. *Tuenti* is a popular OSN in Spain. The results of the tool were high in accuracy.

El Azab *et al.* [14], proposed a classification method for fake accounts detection on Twitter. This method aims to detect fake accounts based on the minimum set of attributes. Initially, determined the main factors, which influence the correct detection of fake accounts. In the next step, the determined factors were used, in order to apply the classification algorithm. The data-set was gathered by the *@FakeProject* and other sources. The fake accounts were bought from three vendors. The results of the study were promising.

The majority of the studies, mentioned above, have focused on the detection of fake accounts (followers). Our study focuses on the detection of fake influencers. Furthermore, the networks of fake followers that have been examined, in the reviewed literature, are not strictly egocentric ones as the links between alters (alter-alter ties) were not crawled. In fact, the authors of the previous studies, take into consideration only

[4]https://networkx.github.io/

[5]https://www.python.org/

[6]https://www.tuenti.com/en/

the profiles of the fake accounts and through these profiles they create a simulation of the Twitter network that is examined. Retrieving ties between alters is a time-consuming task due to the Twitter's API rate limits: 15 account requests per 15 minute interval. Getting advantage of having crawled the full egocentric Twitter network of several legitimate and fake influencer networks we also propose, in this study, several network characterisation measures, in addition to the centrality ones, for the detection of fake influencers on Twitter.

## III. Methodology

To conduct our empirical study we acquired publicly accessible Twitter user egocentric networks. This is a very time-consuming procedure because for every Twitter account (ego) we crawled we had to find all:

- ego's alters (friends and followers)
- alter - alter ties. This means that for every ego's alter we have to find which of its alters (alters of alter) belong to the set of ego's alters

To get an impression of the complexity of this process imagine a Twitter account with 5000 alters each of which has on average 4000 alters. To construct the egocentric network of this account $5000x4000 = 200000$ checks are required. In practice both fake and legitimate influencers have much more than 5000 alters, so we decided to limit our investigation to maximum 5000 alters of each account prioritising to *ego*'s alters that have bidirectional ties with the *ego*.

Given the restrictions posed by the Twitter API (maximum 15 user requests per 15 minutes) and despite the optimisation of our crawling program it takes, on average, 8-10 hours to create the egocentric network of each one of Twitter accounts that were crawled. However, we strongly believe that the dataset we have created it's an important contribution its own to the corresponding research and, therefore, it is freely available online[7] for everyone who wishes to use it either to validate the current work or to perform her/his own.

### A. Dataset and data collection

Our dataset consists of 36 egocentric networks (stored in Pajek[8] format) corresponding to 18 legitimate and 18 fake influencers. The networks were crawled, using the procedure and restrictions mentioned in the previous paragraph, during the period March 2018 - May 2018. We manually selected legitimate influencers among politicians, journalists, TV personas, football players and marketing specialists from a variety of countries including Cyprus, Turkey, Italy, UK and USA.

Finding fake influencers was a bit more tricky. We had first to give a working definition of 'fake influencer'. After a thorough study we ended up to use the following definition: "Fake influencer in Twitter is an account whose a great proportion (higher than 50%) of followers are fake".

The next step was to identify fake influencers. For this purpose we registered a Twitter account (*@andreast88*) and following the methodology of Cresci *et al.* [4] we bought 1000 fake followers from three different vendors. Then, we crawled candidate fake influencers by checking other Twitter accounts that our fake followers follow. The egocentric network of *@andreast88* account is shown in Figure 1. Note that due to the measures taken by Twitter regarding fake accounts, on March 2018, most of *@andreast88* eventually disappeared. However, we had already identified thousands of candidate fake influencers among which 18 were manually selected and crawled through the Twitter API.

We first started by crawling *@andreast88* followers (fake accounts' alters) and constructed the egocentric network of accounts that had high volume of followers. The third dataset consists of legitimate Twitter accounts. We have acquired 1.5 degree ego network for each legitimate and fake Twitter account. Finally an independent samples $t$-test was conducted to compare the average centrality degree of legitimate and fake influencer accounts. We conducted the $t$-test for the centrality measures used by the authors [3] and for the centrality measures we proposed.

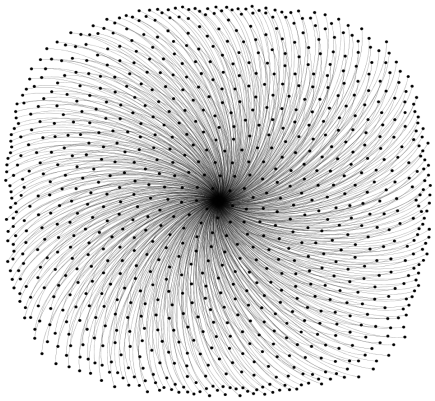

Fig. 1. @andreast88 egocentric network consists of fake followers and has a clearly star structure - visualisation using Frutcherman Reingold algorithm [15].

### B. Mathematical background & suggested measures

Centrality measures have been used by several researchers as a way to identify influencers in social networks since centrality can describe an actor's relative position within the context of their social networks. Influencers, on the other hand, are network actors that are important taking into consideration some criteria (expertise, popularity, prestige, etc).

Mehrotra *et al.* [3] used several centrality measures including betweenness, eigenvector, out-degree, Katz and Load centrality to detect fake accounts building on the idea that these centrality measures are fundamentally related to the concept of social influence. However, as already mentioned they computed these measures in simulations for the real egocentric networks of the fake followers since the alter-alter ties were not included in the network.

[7] https://irci.eu/fake-influencers/
[8] http://mrvar.fdv.uni-lj.si/pajek/

In this work we compare these measures with network characterisation measures such as centralisation, density and reciprocity, along with some other centrality measures not used before (pagerank centrality and authority) using the actual egocentric Twitter networks of real and fake influencers. We explain these measures below while Table I summarises the definitions.

We follow the typical notation of network representation as a graph $\mathcal{G}$, i.e., $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ being the set of vertices (nodes) and $\mathcal{E}$ being the set of edges (ties between vertices).

*1) Relative average degree:* The degree centrality is a typical measure of centrality. However, the average degree centrality is a measure for global network characterisation. Typically fake users have low in degree and as a result one expects that fake influencer egocentric networks will have a rather low average degree centrality. In this work we use a slightly different definition of the average degree centrality as shown in eq. 1. We divide the sum of all node degrees by *ego*'s degree instead of the number of network nodes.

$$\bar{D} = \frac{\sum_{v \in \mathcal{V}} \check{d}_g[v]}{\check{d}_g[v_e]} \tag{1}$$

where $\mathcal{V}$ is the set of network vertices, $\check{d}_g[v]$ indicates the *in degree* of vertex $v$, and $v_e$ is the *ego* node.

*2) Ego's out degree centrality:* The *out degree* of the *ego* in celebrity and fake influencer networks is usually small compared to the *in degree* of the *ego* or the number of vertices in the network. As a result the *out degree centrality* was suggested from some researchers [3] as a measure for identifying fake accounts. We consider that *out degree centrality* could be used for differentiating legitimate users from fake influencers, however, we do not expect that this measure could be used to separate real influencers, such as celebrities and field experts, from fake influencers. Nevertheless, for comparison purposes we also included this measure in our experiments.

$$\hat{D}_C[v_e] = \frac{\hat{d}_g[v_e]}{g-1} \tag{2}$$

where $\hat{d}_g[v_e]$ indicates *ego's out degree* and $g = |\mathcal{V}|$ is the total number of vertices (cardinality of set $\mathcal{V}$) in the network.

*3) Density:* The *density* of a network is a measure of completeness. A network with density equal to one is a fully connected network. Fake influencer networks are expected to have a very low density since the number of alter-alter ties would be minimal. It is very unlikely for fake users to follow each other. The density of a directed graph is given by:

$$d = \frac{e}{g \cdot (g-1)} \tag{3}$$

where $e = |\mathcal{E}|$ is the total number of edges in the network and $g$ is the number of vertices.

Density is affected by the number of vertices in the network; the larger the number of vertices the smaller the density. Thus, to account for this we used a slightly different formula than the one shown in eq. 3:

$$d = \frac{log_2(g)}{2} \cdot \frac{e}{g \cdot (g-1)} \tag{4}$$

*4) Ego's eigenvector centrality:* Eigenvector centrality is another well-known centrality measure that is used to identify users that are dominating a network. The *eigenvector centrality* of a vertex is computed based on the centralities of its neighbours (direct and indirect). Thus, it is expected that provides a more accurate measure of centrality compared to *degree centrality*. Since the *eigenvector centrality* is divided across all network vertices we expect that legitimate users will have lower *eigenvector centrality* than real influencers and the latter lower than fake influencers.

*Eigenvector centrality* was suggested by Mehrotra *et al.* [3] as a measure for identifying fake accounts. For comparison purposes we also included this measure, denoted as $E_C[v_e]$, in our experiments.

*5) Ego's pagerank centrality:* Pagerank centrality is a centrality measure computed with aid of the well known pagerank algorithm[9] originally proposed for the ranking of websites in terms of reliability and prestige. The *pagerank centrality* of a vertex is computed based on the centralities of its direct neighbours; thus it's an iterative algorithm and usually computationally less efficient than eigenvector centrality. However, in terms of centrality estimation pagerank is considered more effective than eigenvector centrality because it is only affected by the immediate neighbours of a vertex. In real-life social networks the influence that a friend of a friend of a friend has to a person is actually negligible. On the other hand this is not the case for computer and other types of networks. For all these reasons we suggest the pagerank centrality as an alternative to eigenvector centrality. The independence to $t$-test shows a higher discrimination (see Table I) ability of the pagerank centrality compared to eigenvector centrality w.r.t real and flake influencer classification.

*6) Average closeness centrality:* The closeness centrality of a vertex is the inverse of the mean geodesic distance of this vertex to the remaining vertices of the network. Formally, is computed with the aid of the formulas 5 and 6:

$$A_D[v] = \frac{\sum_{j=1, j \neq v}^{g} d(v,j)}{g-1} \tag{5}$$

$$C_C[v] = \frac{1}{A_D[v]} \tag{6}$$

where $d(v,j)$ is the geodesic distance between vertices $v$ and $j$, $A_D[v]$ denotes the mean geodesic distance of vertex $v$ to the remaining vertices of the network while $C_C[v]$ is the closeness centrality of vertex $v$.

In egocentric networks the *ego* node is connected to every other vertex either through an incoming edge (followers), outgoing edge (friends) or bidirectional edge (both friends and followers). As a result the closeness centrality $C_C[v_e]$ of the ego may not differ in the networks of real and fake influencers.

[9]https://en.wikipedia.org/wiki/PageRank

In contrary, the average closeness centrality, denoted as $\bar{C}_C$ and given by eq. 7, is expected to be higher in legitimate influencer networks compared to that of fake influencers. The rational is simple: ties between fake followers are very unlikely while fake followers have very few (if any) incoming edges, so they are practically non-reachable from other network vertices.

We should note here that by averaging the closeness centralities of all network vertices we practically create a measure of compactness of the network. That's why we mention the type of $\bar{C}_C$ in Table I as global network.

$$\bar{C}_C = \frac{\sum_{v \in \mathcal{V}} C_C[v]}{g} \tag{7}$$

*7) Ego's betweenness centrality:* The betweenness centrality of a vertex $v$ is the proportion of shortest paths between any two other vertices of the network that pass through $v$. Betweenness centrality is one of the measures proposed by Mehrotra *et al.* [3] to detect fake accounts and it was included in the current study for comparison purposes. In egocentric networks the betweenness centrality of the ego indicates the absence of alter-alter ties. However, as already mentioned previously in fake influencer networks the majority of fake followers are isolated since they do not have incoming edges. As a result very few pairs of fake followers reach each other and the betweenness centrality measure becomes misleading. In order to avoid this problem we apply betweenness centrality on the undirected graph corresponding to the egocentric network of user. In this case all vertices communicate, either directly (alter -alter ties) or through *ego*. We expect that the betweenness centrality of the ego in fake influencer networks will be higher than its counterpart of legitimate influencer networks.

Mehrotra *et al.* [3] used, also, in their work the *Load centrality*. Through our experimentation it appears that betweenness and Load centrality have a very high degree of correlation, thus, there is no reason to include both of them in our investigation.

*8) Ego's reciprocity of incoming edges:* Reciprocity is a measure of the likelihood of vertices in a directed network to be mutually linked. Low reciprocity indicates a hierarchically structured network while high reciprocity corresponds a network of peers. The reciprocity of ego's incoming edges indicates whether the *ego* follows back her/his followers. Ordinary Twitter users have, usually, high reciprocity of their incoming ties. Legitimate influencers have low reciprocity of their incoming ties but the corresponding value of fake influencers is expected to be even lower: Ego would never follow back fake followers. Ego's reciprocity of incoming edges is computed with the aid of eq. 8:

$$R[v_e] = \frac{log_2|\mathcal{N}_{\mathcal{O}}[v_e]|}{log_2|\mathcal{N}_{\mathcal{I}}[v_e]|} \tag{8}$$

where $\mathcal{N}_{\mathcal{I}}[v_e]$ is the set of vertices that are connected to ego ($v_e$) with an incoming edge (followers), $\mathcal{N}_{\mathcal{O}}[v_e]$ is the set of vertices that are connected to ego with an outgoing edge (friends) and $|\cdot|$ indicates the cardinality of a set. The logarithm in eq. 8 is used to account for the fact that in both legitimate and fake influencers the proportion of friends to followers is very low.

*9) Centralisation:* Centralisation can indicate the extent to which the network is dominated by one node. It is a very effective network characterisation measure, especially for egocentric networks. A network with the highest centralisation (value equal to one) has a star structure while networks with centralisation equal to zero are complete networks. Centralisation for directed graphs is computed with the aid of eq. 9:

$$C_e = \frac{g \cdot d_{max} - e}{e \cdot (g-1)} \tag{9}$$

where $d_{max}$ is the highest in-degree in the network (in egocentric networks this is typically the in degree of *ego*).

*10) Ego's Authority:* The authority of a person expresses the degree to which is respected by knowledgeable people in the community. Authority is a classic measure of centrality originally proposed to evaluate the prestige of a website. It is typically computed with the aid of HITS (Hyperlink-Induced Topic Search) algorithm. HITS is an iterative algorithm that computes two scores per document (here vertex): hub and authority. Zhang *et al.* [16] used HITS algorithm in order to detect experts in a closed domain while Giannoulakis *et al.* [17] used it crowdsourcing image annotation environments to identify effective annotators (high hub value) and descriptive image tags (high authority value).

The authority value of a network node depends heavily on the number of incoming edges of this node and it is shared among all network nodes. As a result we expect that ordinary Twitter users will have a low authority value since some celebrities - influencers should be present in their egocentric network while legitimate influencers would have a higher authority value. However, fake influencers would have even higher authority value because it is very unlikely that any other celebrity - influencer would be present among their (fake influencers') followers while celebrities - influencers within their (fake influencers') friends it is unlikely to share with them the same fake followers.

The *Python* code below shows how the previously mentioned measures (summarised in the first column of Table I) were computed with the aid of the *networkx* library:

```
>>> import numpy as np
>>> import networkx as nx
>>> import operator
>>> Q = nx.read_pajek(filepath)
>>> Q = nx.DiGraph(Q)
>>> Q1 = Q.to_undirected(reciprocal=False)
>>> e = Q.number_of_edges()
>>> e1 = Q1.number_of_edges()
>>> g = Q.number_of_nodes()
>>> sorted_d = sorted(Q.in_degree().items(),
    key=operator.itemgetter(1),reverse=True)
>>> ego = sorted_d[0][0]
>>> ego_degree = sorted_d[0][1]
>>> L = list(Q.in_degree().values())
>>> D = np.mean(L)/ego_degree
>>> D_C = Q.out_degree()[ego]/(g-1)
>>> d = nx.density(Q)*np.log2(g)/2
>>> E = nx.eigenvector_centrality(Q, max_iter=1000,
    tol=1e-06, nstart=None)
>>> E_C = E[ego]
```

TABLE I
MEASURES AND SIGNIFICANCE OF RESULTS

| Measure | Type | Definition | $t$-value | $p$ value $\leqslant$ | significant ($p < 0.01$) |
|---|---|---|---|---|---|
| Relative average degree ($\bar{D}$) | global network | suggested measure, see eq. 1 | 7.3503 | 0.00001 | YES |
| Out degree centrality ($\hat{D}_C[v_e]$) | centrality | measure proposed in [3] | 4.0801 | 0.00026 | YES |
| Density ($d$) | global network | suggested measure, see eq. 4 | 7.2963 | 0.00001 | YES |
| Eigenvector centrality ($E_C[v_e]$) | centrality | measure proposed in [3] | $-3.1695$ | 0.00323 | YES |
| Pagerank centrality ($P_C[v_e]$) | centrality | suggested measure, alt. to eigenvector centrality | $-6.9555$ | 0.00001 | YES |
| Average closeness centrality ($\bar{C}_C$) | global network | suggested measure, see eq. 6 | 16.3864 | 0.00001 | YES |
| Betweenness Centrality ($B_C[v_e]$) | centrality | measure proposed in [3] (modified, see Sec. III-B7) | $-12.3484$ | 0.00001 | YES |
| Reciprocity ($R[v_e]$) | centrality | suggested measure, see eq. 8 | 6.0032 | 0.00001 | YES |
| Centralisation ($C_e$) | global network | suggested measure, see eq. **??** | $-5.5483$ | 0.00001 | YES |
| Authority ($A[v_e]$) | centrality | suggested measure, see also [17] | $-5.4705$ | 0.00001 | YES |

```
>>> P = nx.pagerank(Q)
>>> P_C = P[ego]
>>> C = nx.closeness_centrality(Q)
>>> C_C = np.mean(list(C.values()))
>>> B = nx.betweenness_centrality(Q1)
>>> B_C = B[ego]
>>> ego_fo = [u for (u,v) in Q.edges() if v==ego]
>>> ego_bi = [u for u in ego_fo if u in Q[ego]]
>>> R = np.log2(len(ego_bi))/np.log2(len(ego_fo))
>>> d_max = max(L)
>>> C_e = (g*d_max-e)/(e*(g-1))
>>> [H,A]=nx.hits(Q)
>>> A_C= A[ego]
```

## IV. RESULTS AND DISCUSSION

The values of all measures described in the previous paragraph for all the 36 egocentric networks investigated are presented in Tables II (fake influencers) and III (legitimate influencers). The corresponding significance scores of the $t$-test are shown in Table I. In the latter we see that all the measures that were investigated differ significantly, at semantic level $a = 0.01$, between legitimate and fake influencers. However, the $p$ value varies across them with the poorest performance recorded for the out degree and eigenvector centralities. In contrary, all the network characterisation measures (denoted as global network in Table I), show excellent discrimination ability reaching $p$ values lower that $0.00001$. For instance, we see in Tables II and III that the lowest value of centralisation $C_e$ in the fake follower accounts ($0.0590$) is higher than the highest ($0.0303$) in legitimate influencers. This means that these two types of accounts are linearly separable even with this single measure. Similar are the cases of relative average degree $\bar{D}$ and density $d$.

The most important finding of this study is probably the fact that the relative average degree $\bar{D}$ measure can be used to detect fake influencers. This measure is the only one in Table I that can be estimated through the alters' profiles and without the need to crawl all alter to alter ties.

## V. CONCLUSION & FURTHER WORK

This paper reports a comparative study of measures that can be used to detect fake influencers on Twitter. This is probably the first time this problem is investigated. The results show that network characterisation measures are equivalently (if not more) effective as the centrality measures for fake influencer detection. Some of the newly proposed measures, such as the centralisation of a network, can be used, on their own, to separate fake from legitimate influencers even linearly. An important outcome of this study is the creation of the first ever dataset of egocentric networks of fake and legitimate influencer Twitter accounts. This dataset is freely available to the research community.

In the near future we will continue collecting egocentric networks of legitimate and fake influencers in order to enlarge, as much as possible, our dataset. In addition, profile characteristics of ego and alters will be investigated for the task of fake influencers detection.

## ACKNOWLEDGMENT

This work is partially funded by the EC project "ENCASE: EnhaNcing seCurity And privacy in the Social wEb: a user centered approach for the protection of minors" under the contract H2020-MSCA-RISE-2015-691025.

## REFERENCES

[1] S. Neti, "Social media and its role in marketing," *International Journal of Enterprise Computing and Business Systems*, vol. 1, no. 2, pp. 1–15, 2011.

[2] S. Gurajala, J. S. White, B. Hudson, B. R. Voter, and J. N. Matthews, "Profile characteristics of fake twitter accounts," *Big Data & Society*, vol. 3, no. 2, 2016.

[3] A. Mehrotra, M. Sarreddy, and S. Singh, "Detection of fake twitter followers using graph centrality measures," in *Proceedings of the 2nd International Conference on Contemporary Computing and Informatics*, Dec 2016, pp. 499–504.

[4] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "A fake follower story: improving fake accounts detection on twitter," IIT-CNR, Tech. Rep. TR-03, 2014.

[5] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, p. 35, mar 1977.

[6] B. Coralio, C. Antoni, and Z. Yves, "Who's who in networks. wanted: The key player," *Econometrica*, vol. 74, no. 5, 2006.

[7] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 6, pp. 811–824, Nov 2012.

[8] X. Zheng, Z. Zeng, Z. Chen, Y. Yu, and C. Rong, "Detecting spammers on social networks," *Neurocomputing*, vol. 159, pp. 27 – 34, 2015.

TABLE II
DETAILED MEASURES - FAKE INFLUENCERS

| Account | # nodes ($g$) | $\bar{D}$ | $\hat{D}_C[v_e]$ | $d$ | $E_C[v_e]$ | $P_C[v_e]$ | $\bar{C}_C$ | $B_C[v_e]$ | $R[v_e]$ | $C_e$ | $A[v_e]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| F01 | 628 | 0.0039 | 0.1962 | 0.0157 | 0.4570 | 0.2209 | 0.0968 | 0.9929 | 0.2030 | 0.4079 | 0.3990 |
| F02 | 644 | 0.0258 | 0.4323 | 0.0765 | 0.0446 | 0.0744 | 0.2123 | 0.9146 | 0.4591 | 0.0590 | 0.0093 |
| F03 | 924 | 0.0100 | 0.1788 | 0.0460 | 0.2176 | 0.1061 | 0.1614 | 0.9260 | 0.3660 | 0.1073 | 0.0338 |
| F04 | 665 | 0.0062 | 0.1536 | 0.0270 | 0.2492 | 0.1447 | 0.0839 | 0.9850 | 0.5010 | 0.2410 | 0.0825 |
| F05 | 1983 | 0.0018 | 0.1065 | 0.0090 | 0.0144 | 0.0865 | 0.0522 | 0.7826 | 0.1387 | 0.2779 | 0.3991 |
| F06 | 1994 | 0.0005 | 0.0000 | 0.0028 | 1.0000 | 0.4599 | 0.0005 | 1.0000 | 0.0000 | 0.9990 | 0.9990 |
| F07 | 4852 | 0.0003 | 0.0368 | 0.0017 | 0.0025 | 0.1449 | 0.0186 | 0.9996 | 0.0979 | 0.7275 | 0.9844 |
| F08 | 1066 | 0.0145 | 0.0310 | 0.0705 | 0.4255 | 0.0838 | 0.1465 | 0.5832 | 0.0472 | 0.0640 | 0.0393 |
| F09 | 1464 | 0.0008 | 0.0212 | 0.0043 | 0.5408 | 0.2888 | 0.0151 | 0.9993 | 0.0428 | 0.8305 | 0.8454 |
| F10 | 4857 | 0.0003 | 0.0290 | 0.0016 | 0.1148 | 0.1802 | 0.0153 | 0.9996 | 0.0712 | 0.7573 | 0.8608 |
| F11 | 2519 | 0.0004 | 0.0087 | 0.0046 | 0.4135 | 0.2790 | 0.0051 | 1.0000 | 0.0053 | 0.9770 | 0.9900 |
| F12 | 3058 | 0.0007 | 0.0697 | 0.0075 | 0.3334 | 0.1729 | 0.0422 | 0.9946 | 0.1802 | 0.4849 | 0.5238 |
| F13 | 888 | 0.0011 | 0.0045 | 0.0055 | 0.0000 | 0.3111 | 0.0031 | 1.0000 | 0.0000 | 0.9944 | 1.0000 |
| F14 | 691 | 0.0141 | 0.3464 | 0.0469 | 0.0249 | 0.1118 | 0.1743 | 0.9478 | 0.2643 | 0.1015 | 0.0074 |
| F15 | 1685 | 0.0061 | 0.1977 | 0.0600 | 0.2316 | 0.0987 | 0.1499 | 0.9332 | 0.4087 | 0.0972 | 0.0499 |
| F16 | 4696 | 0.0017 | 0.2863 | 0.0074 | 0.0054 | 0.1054 | 0.1253 | 0.9882 | 0.2561 | 0.1252 | 0.7481 |
| F17 | 911 | 0.0015 | 0.1714 | 0.0064 | 0.2574 | 0.2150 | 0.0766 | 0.9978 | 0.0405 | 0.7154 | 0.9373 |
| F18 | 3552 | 0.0003 | 0.0008 | 0.0033 | 0.5316 | 0.3075 | 0.0006 | 1.0000 | 0.0017 | 0.9978 | 0.9986 |

TABLE III
DETAILED MEASURES - LEGITIMATE INFLUENCERS

| Account | # nodes ($g$) | $\bar{D}$ | $\hat{D}_C[v_e]$ | $d$ | $E_C[v_e]$ | $P_C[v_e]$ | $\bar{C}_C$ | $B_C[v_e]$ | $R[v_e]$ | $C_e$ | $A[v_e]$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| L01 | 1200 | 0.1068 | 0.3528 | 0.5007 | 0.0891 | 0.0133 | 0.4496 | 0.2270 | 0.5443 | 0.0070 | 0.0046 |
| L02 | 1825 | 0.0791 | 0.2944 | 0.3889 | 0.0843 | 0.0071 | 0.4640 | 0.1343 | 0.4650 | 0.0066 | 0.0045 |
| L03 | 3433 | 0.0476 | 0.4053 | 0.2408 | 0.0787 | 0.0081 | 0.4442 | 0.2191 | 0.5592 | 0.0059 | 0.0031 |
| L04 | 2183 | 0.0560 | 0.4500 | 0.4298 | 0.0950 | 0.0104 | 0.4289 | 0.2594 | 0.4742 | 0.0077 | 0.0046 |
| L05 | 975 | 0.0760 | 0.2936 | 0.3189 | 0.0903 | 0.0176 | 0.3475 | 0.3643 | 0.4044 | 0.0125 | 0.0700 |
| L06 | 1628 | 0.0484 | 0.8679 | 0.1352 | 0.1501 | 0.0203 | 0.3920 | 0.6652 | 1.0071 | 0.0121 | 0.0059 |
| L07 | 718 | 0.1626 | 0.2455 | 0.7091 | 0.0875 | 0.0083 | 0.4970 | 0.0825 | 0.4925 | 0.0072 | 0.0062 |
| L08 | 1660 | 0.0488 | 0.3279 | 0.2151 | 0.0849 | 0.0113 | 0.3491 | 0.3096 | 0.3548 | 0.0117 | 0.0063 |
| L09 | 585 | 0.0536 | 0.4144 | 0.2192 | 0.1946 | 0.0392 | 0.3514 | 0.5534 | 0.4796 | 0.0303 | 0.0170 |
| L10 | 1919 | 0.0823 | 0.0537 | 0.4382 | 0.1076 | 0.0233 | 0.4193 | 0.3922 | 1.4281 | 0.0058 | 0.0035 |
| L11 | 764 | 0.1993 | 0.6387 | 0.8429 | 0.0885 | 0.0070 | 0.5260 | 0.1176 | 0.7039 | 0.0053 | 0.0047 |
| L12 | 625 | 0.1396 | 0.7228 | 0.2463 | 0.0525 | 0.0155 | 0.3631 | 0.6032 | 0.5337 | 0.0099 | 0.0050 |
| L13 | 1059 | 0.0605 | 0.2618 | 0.2984 | 0.1397 | 0.0287 | 0.4040 | 0.4102 | 0.5816 | 0.0147 | 0.0078 |
| L14 | 619 | 0.1376 | 0.6456 | 0.3459 | 0.0266 | 0.0250 | 0.3862 | 0.5945 | 0.6182 | 0.0101 | 0.0029 |
| L15 | 2177 | 0.0422 | 0.1691 | 0.1169 | 0.0922 | 0.0107 | 0.3586 | 0.2674 | 0.2774 | 0.0104 | 0.0072 |
| L16 | 4194 | 0.0276 | 0.0596 | 0.1639 | 0.0920 | 0.0077 | 0.4130 | 0.1104 | 0.4198 | 0.0084 | 0.0047 |
| L17 | 1401 | 0.1053 | 0.2893 | 0.5058 | 0.0881 | 0.0057 | 0.4867 | 0.0642 | 0.5230 | 0.0061 | 0.0046 |
| L18 | 2183 | 0.0560 | 0.2227 | 0.2779 | 0.0950 | 0.0104 | 0.4289 | 0.2594 | 0.4742 | 0.0077 | 0.0046 |


[9] Y. Shen, J. Yu, K. Dong, and K. Nan, "Automatic fake followers detection in chinese micro-blogging system," in *Advances in Knowledge Discovery and Data Mining*, V. S. Tseng, T. B. Ho, Z.-H. Zhou, A. L. P. Chen, and H.-Y. Kao, Eds. Cham: Springer International Publishing, 2014, pp. 596–607.

[10] Y. Zhang and J. Lu, "Discover millions of fake followers in weibo," *Social Network Analysis and Mining*, vol. 6, no. 1, Mar 2016.

[11] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?" in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, Aug 2014, pp. 620–627.

[12] 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*, ser. IMC'13. New York, NY, USA: ACM, 2013, pp. 163–176.

[13] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*. USENIX Association, 2012, pp. 15–15.

[14] A. E. Azab, A. M. Idrees, M. A. Mahmoud, and H. Hefny, "Fake account detection in twitter based on minimum weighted feature set," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 10, no. 1, pp. 13 – 18, 2016.

[15] P. Gajdos, T. Jezowicz, V. Uher, and P. Dohnálek, "A parallel fruchterman-reingold algorithm optimized for fast visualization of large graphs and swarms of data," *Swarm and Evolutionary Computation*, vol. 26, pp. 56–63, 2016.

[16] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: Structure and algorithms," in *Proceedings of the 16th Int'l Conference on World Wide Web*, ser. WWW '07. New York, NY, USA: ACM, 2007, pp. 221–230.

[17] S. Giannoulakis, N. Tsapatsoulis, and K. Ntalianis, "Identifying image tags from instagram hashtags using the hits algorithm," in *Proceedings of the 2017 IEEE Cyber Science and Technology Congress*, Nov 2017, pp. 89–94.