On Using Node Indices and Their Correlations for Fake Account Detection

Sara Asghari
Department of Computer Engineering
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
sara0asghari@gmail.com

Mostafa Haghir Chehreghani
Department of Computer Engineering
Amirkabir University of Technology (Tehran Polytechnic)
Tehran, Iran
mostafa.chehreghani@aut.ac.ir

Morteza Haghir Chehreghani
Department of Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden
morteza.chehreghani@chalmers.se

***Abstract*—With the growing rate of online social networks, the number of fake accounts is multiplying day by day. There exist many approaches in the literature that try to distinguish fake accounts from real ones, for example, those that use machine learning and classification techniques to learn whether a user should be labeled as fake (bot) or not. In this paper, we follow a different approach and try to use node measurements in the field of *complex networks analysis* to identify fake accounts. We first model users' interactions with a large graph. For example, in Twitter, we can form graphs of follower-following, comments, retweets, mentions, and so on. We then investigate different measurements, such as centrality indices and their correlations, to separate real and fake accounts. We find that measurements such as *average path length*, *eigenvector* centrality, *harmonic* centrality, *degree*, *local reaching* centrality and their correlations provide good indicators to distinguish real and fake accounts.**

***Index Terms*—Online social networks, Twitter, fake account detection, complex networks analysis, centrality measures**

I. Introduction

In recent years, due to low price, high accessibility, and fast information propagation, online social networks such as Twitter have become one of the main news resources for millions of people throughout the world. With the daily growth of these networks, the number of fake accounts that are generated to propagate false news and misinformation increases, too. Twitter in 2014 announced that between 5 and 8.5 percent of its accounts are fake accounts (bots). In another estimation, Varol et al. [20] estimated that the number of such accounts is between 9 and 15 percent. By disseminating false news and misinformation, fake accounts can affect the decisions of millions of real users.

There exist many approaches in the literature that try to distinguish fake accounts from real ones. Two main categories of existing methods include *graph-based models* that use the assumption that in a social network, bots and human users behave differently in terms of network connectivity and community membership; and *feature-based models* that try to extract features discriminating fake accounts from real ones and exploit these features in a classification algorithm, such as random forest, neural networks, logistic regression, and support vector machines (SVM), to classify accounts into fake and real ones. A positive aspect of graph-based methods is that they usually do not require a large amount of labeled training data.

Existing graph-based methods consider only simple node measurements and indices, such as *degree* [4] and *random walks* [14], to separate fake and real accounts. However, in the field of complex networks analysis, there exist many other measurements that are used to assign scores to nodes. Various measurements are based on different roles that nodes find in the network. Hence, they reflect different structural properties of nodes in the network. The applicability and usefulness of these indices are well-studied in different domains such as urban street networks [17], academic networks [19], and biological networks [16]. However, most of them are not studied for fake account detection yet. Moreover, correlations between different indices can give insightful information about real and fake accounts. Such correlations are not extensively investigated for fake account detection.

In this paper, we go beyond simple measurements and examine several other node indices and their correlations to distinguish fake and real accounts. First, using relations such as follower-following, comments, retweets, and mentions, we model users' interactions with a large graph. Then, we examine several node indices such as *average path length*, *eigenvector* centrality, *harmonic* centrality, *degree*, *local reaching* centrality, and notice that real and fake accounts find different values for them. Hence, they can be used to distinguish these two types of accounts. Finally, we notice that correlations between a number of node indices, namely *average shortest path length* and *harmonic* centrality, *eigenvector* centrality and *in-degree*, *eigenvector* centrality and *PageRank*, *out-degree* and *in-degree*, *out-degree* and *PageRank*, *harmonic* centrality and *eigenvector* centrality, can be used to detect bots, as real and fake accounts show different correlations for these pairs of indices.

978-1-6654-8045-1/22/$31.00 ©2022 IEEE

The rest of this paper is organized as follows. In Section II, we present a brief overview of related work. In Section III, we describe our proposed method for distinguishing fake and real accounts. In Section IV, we present the details of our experiments and the obtained results. Finally, the paper is concluded in Section V.

II. Related work

Two main categories of bot detection algorithms include graph-based methods and feature-based approaches. In the following, we briefly describe some known methods from each category.

a) Graph-based methods: Graph-based bot detection methods are based on the assumption that in a social network, bots and human users behave differently in terms of network connectivity and community engagement. Cao et al. [5] present the *SybilRank* algorithm. In this model, it is assumed that fake accounts (Sybils) in online social networks have a considerably smaller number of connections to real accounts (non-Sybils). Moreover, they are mostly connected to other Sybils. In the SybilRank algorithm, accounts are ranked by their Sybil-likelihood score, which is computed based on the social network properties and the landing probability of short random walks. Breuer, Eilat, and Weinsberg [3] present the *SybilEdge* algorithm, which detects new fake accounts by leveraging two individual-level differences in how new fake accounts interact with other accounts: (i) their choices of friend request targets, and (ii) how the selected targets respond to the new fakes' friend requests.

b) Feature-based methods: Varol et al. [20] introduce a machine learning system that extracts more than a thousand features in six different classes: users and friends metadata, tweet content and sentiment, network patterns, and activity time series. Their analysis shows that user metadata and content features are the two most informative data sources for detecting simple bots. They evaluate their system using several classification algorithms, among which random forest obtains the best classification performance. Kudugunta and Ferrara [13] present a deep neural network based on contextual long short-term memory (LSTM) architecture that exploits both tweet content and metadata to detect bots at the tweet level. They use only a small number of features to construct an interpretable model that can be trained faster and is less prone to overfitting.

Bharti and Pandey [1] propose a two-phase model (feature selection phase and classification phase) to deal with fake account detection. They apply techniques such as information gain, correlation, and minimum redundancy maximum relevance, to select an informative subset of features from the original feature space. Then, they use a logistic regression classifier for fake and real account classification. Homsi et al. [12] study the effect of two reduction techniques (PCA and correlation) along with four classifier algorithms (J48, Random Forest, KNN, and Naive Bayes). Their results show that the random forest algorithm, along with correlation data reduction, gives the best accuracy.

III. Our proposed method

In every social network, users interact with each other in the context of several social graphs. Each node represents a user in each of these graphs, and each edge represents a relationship between two users. A ratio of nodes are fake accounts, and the rest are real. Accordingly, several graphs are formed that can be analyzed using a collection of tools and criteria. In this paper, with the help of profile information and tweet metadata in our datasets, we build a *follower-following* graph and a *comments* graph. Follower-following graph is an unweighted social graph in which an edge from node A to node B indicates that user A has followed user B. Comments graph is a weighted social graph in which an edge from node A to node B demonstrates that user B has commented on user A's tweets, and the number of comments indicates the edge weight. Figure 1 presents visualizations of these two graphs, wherein green dots present real accounts, and red dots present fake accounts.

In this paper, we aim to analyze graphs of real and fake Twitter accounts, in order to explore the behavior of bots compared to real humans. This analysis is done by means of node measurements and indices widely used in complex networks analysis, including *in-degree*, *out-degree*, *betweenness*, *harmonic*, *eigenvector*, *Katz*, *PageRank*, *clustering coefficient*, *average shortest path length*, *average neighbor degree* and *local reaching*.

For a node v, its *in-degree* centrality is defined as the fraction of nodes its incoming edges are connected to; and its *out-degree* centrality is defined as the fraction of nodes its outgoing edges are connected to [18]. *Betweenness* centrality measures how frequently a node lies on the shortest paths between other nodes. It is defined as follows:

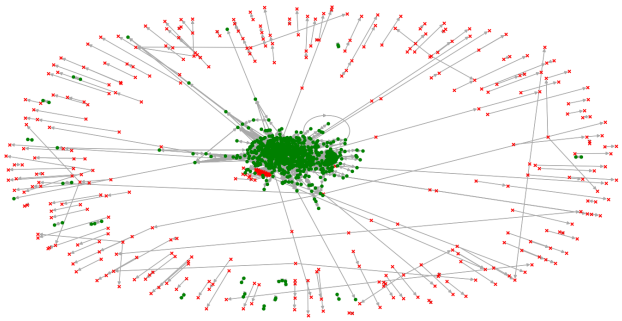
$$c_b(v) = \sum_{s,t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

where V is the set of nodes of the graph, $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through v and σ_{st} is the total number of shortest paths from s to t [2], [6], [9]. *Harmonic* centrality of a node v is the sum of the reciprocal of the shortest path distances from all other nodes to v [18]. It can be applied to both disconnected and connected graphs. *Eigenvector* centrality of node v is recursively defined as follows:

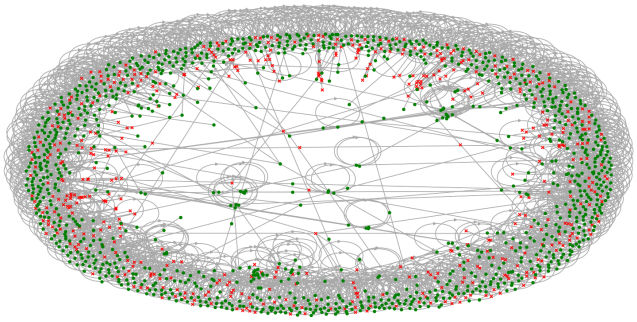
$$c_e(v) = \frac{1}{\lambda} \sum_{u \in N(v)} c_e(u),$$

where $N(v)$ and λ are respectively the set of neighbors of v and some fixed constant [18].

Katz centrality of a node v is computed as the number of its neighbors and all other nodes in the network that connect to v through its neighbors. However, the number of nodes that are connected to v via a path of length d, is penalized by α^d, where α is a real constant between 0 and 1 [18]. In *PageRank* centrality, the score of a node is divided by the number of its



(a) follower-following social graph



(b) comments social graph

Fig. 1: Green solid "O"s present real accounts, and red "X"s represent fake accounts. Nodes with zero degrees have been removed.

outgoing links such that each neighbor gets an equal fraction of the node's score. More formally, it is defined as follows:

$$c_p(v) = \frac{1}{\lambda} \sum_{u \in IN(v)} \frac{c_p(u)}{deg^{out}(u)} + \beta,$$

where $IN(v)$ is the set of in-neighbors of v, $deg^{out}(u)$ is the *out-degree* of node u, λ is a constant that determines the effect of *PageRank* scores and β is a bias term that avoids zero centrality values [7], [18].

Local reaching centrality of a node v in a directed graph is the proportion of the nodes of the graph that are reachable from v [15]. *Clustering coefficient* of a node determines how well-connected its neighbors are and hence, how close they are to forming a complete subgraph [18]. This index for a node v is computed as the proportion of the number of edges between the neighbors of v divided by the maximum number of edges that may exist between them [18]. The *average shortest path length* determines the average distance of each node from all other nodes in the network [8], [18].

Over the constructed graphs, we examine the values of these indices for real and fake accounts (nodes). We then check if there exist considerable differences between these values for real and fake accounts. Furthermore, there usually exist correlations between the values that a pair of these indices find. We investigate such correlations between all pairs of the indices to extract meaningful and discriminating correlation information.

Dataset	Group name	#accounts	#tweets
Cresci-2015	E13	1,481	2,068,037
	FSF	1,169	22,910
Cresci-2017	Genuine accounts	3,474	8,377,522
	Social spambots #2	3,457	428,542

TABLE I: Breakdown of the used datasets.

IV. Experimental Results

In this section, we empirically evaluate the ability of the discussed indices and their correlations, in separating fake and real accounts.

A. Datasets

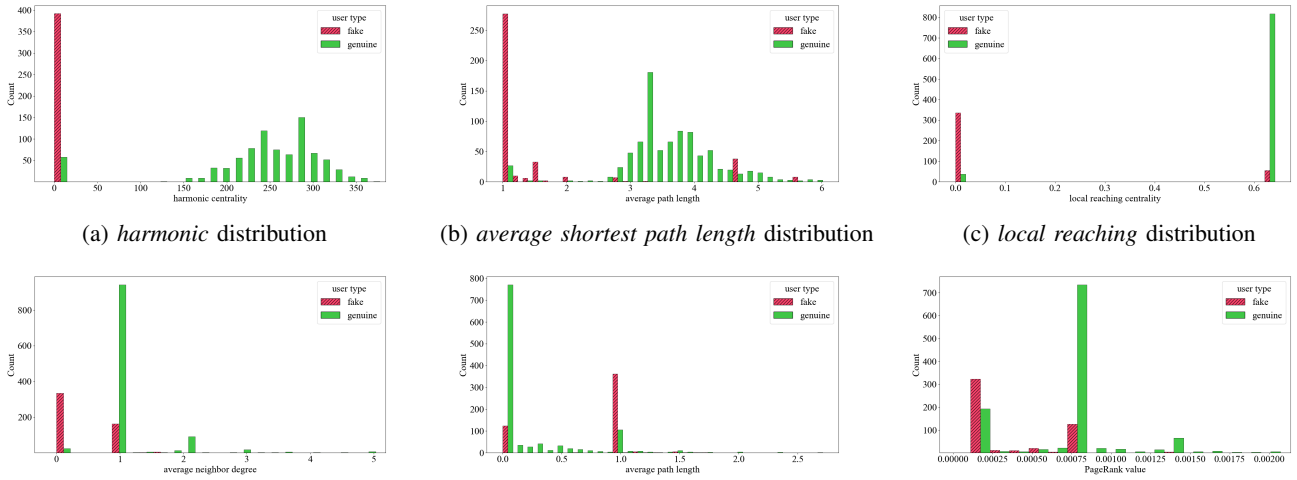
We use two *MIB* Twitter datasets: *Cresci-2015* [10] and *Cresci-2017* [11][1]. Cresci-2015 contains user profile details, tweet metadata and tweets, and follower and following IDs. A mixture of its `E13` and `FSF` sets are used for building the follower-following graph. Likewise, Cresci-2017 includes user profile details, tweet metadata, and tweets. A mixture of its `genuine accounts` and `social spambots #2` sets are used for building the comments graph. The characteristics of the datasets are summarized in Table I. Although at first the number of real and fake nodes is almost equal in both graphs, in the pre-processing step all zero-degree nodes (most of which are fake) are removed.

B. Results

First, for each index, we determine its value for each node and build a distribution histogram, wherein the horizontal axis shows different values that nodes find, and for each value the vertical axis presents the number of real nodes (in green bars) and the number of fake nodes (in red hatched bars) that have this value. Among all the histograms, in the histograms of the following indices, real and fake accounts show more distinction: *harmonic* centrality, *average shortest path length*, and *local reaching* centrality in the follower-following graph, and *average neighbor degree*, *average shortest path length*, and *pagerank* in the comments graph. We depict these histograms in Figure 2. When calculating the value of *harmonic* centrality and *average shortest path length*, the distance from one node to all other nodes is calculated. Since most of the fake nodes are located in small unconnected subgraphs, the value of these two indices is significantly lower for them than for real nodes. Also, *local reaching* of a node considers all nodes that are directly or indirectly connected to it. It can be argued that since fake accounts are usually reachable from much fewer nodes, they have a smaller *local reaching* centrality than real nodes. In our experiments, some indices such as *betweenness* centrality and *clustering coefficient* do not provide a nice distinction between real and fake accounts. Therefore, we do not bring their histograms.

Next, we study the correlations between different pairs of node indices for real and fake accounts. The charts of those

[1]The datasets are publicly available at: https://botometer.osome.iu.edu/bot-repository/datasets.html



(a) *harmonic* distribution (b) *average shortest path length* distribution (c) *local reaching* distribution

(d) *average neighbor degree* distribution (e) *average shortest path length* distribution (f) *pagerank* distribution

Fig. 2: Green bars present the distribution of real users, and red hatched bars depict the distribution of fake users. Histograms 2a, 2b, and 2c are taken from the follower-following graph and histograms 2d, 2e, and 2f are taken from the comments graph. Outliers are removed.

correlations that contain the most valuable and meaningful information about the behaviors of real and fake accounts are presented in Figure 3. Regarding the correlation between *in-degree* and *out-degree*, we can interpret the observed trend as that a real account who follows more individuals (higher *out-degree*) is usually followed by more individuals as well (higher *in-degree*). Nevertheless, a fake account who follows more individuals (higher *out-degree*) may be followed by fewer individuals (lower *in-degree*). The correlation between *out-degree* and *PageRank* displays a similar trend. This is because *PageRank*, similar to *in-degree*, calculates the score of each node based on the scores of its incoming edges.

It is worth mentioning that in the correlation between *average shortest path length* and *harmonic centrality*, real nodes' behavior is natural and logical. However, fake nodes behave abnormally. To be precise, an increase in the *average shortest path length* value means moving away from the center of the graph. Thus, the value of *harmonic centrality*, which is inversely proportional to the distance from the center, must decrease. However, the *harmonic* value is almost zero in fake nodes and remains nearly constant as the value of *average shortest path length* increases.

V. Conclusion

In this paper, a complex networks analysis approach was taken to distinguish real and fake accounts of the Twitter social network. Two social graphs were created, a follower-following graph and a comments graph, in which a small ratio of nodes are fake accounts and the rest are real. Then, node indices used in analyzing complex networks are exploited to study the behavior of fake and real accounts. Among the studied indices, *harmonic* centrality, *average shortest path length*, *local reaching*, and several correlations between different nodes' indices revealed good performance in distinguishing real nodes from fake ones.

The results of this paper and the found differences between the behaviors of real and fake accounts can be used to design better machine learning techniques to identify fake nodes. Strictly speaking, in many machine learning methods, a feature vector is constructed from measurable characteristics extracted from the user profile and user activities. In order to amplify this vector, each of the studied criteria can be added as a new feature to this vector, which, in the end, may result in a more robust classification algorithm.

References

[1] Kusum Kumari Bharti and Shivanjali Pandey. Fake account detection in twitter using logistic regression with particle swarm optimization. *Soft Computing*, 25(16):11333–11345, 2021.

[2] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[3] Adam Breuer, Roee Eilat, and Udi Weinsberg. Friend or faux: graph-based early detection of fake accounts on social networks. In *Proceedings of The Web Conference 2020*, pages 1287–1297, 2020.

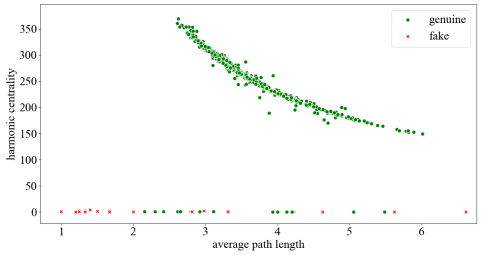
[4] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors, *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, pages 665–674. ACM, 2011.

[5] Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. Aiding the detection of fake accounts in large scale social online services. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pages 197–210, 2012.

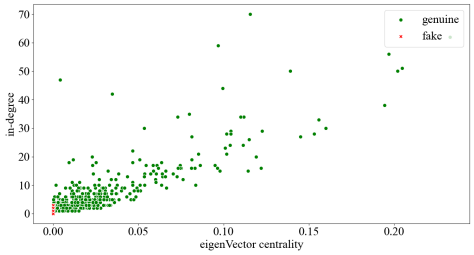
[6] Mostafa Haghir Chehreghani. An efficient algorithm for approximate betweenness centrality computation. *Comput. J.*, 57(9):1371–1382, 2014.

[7] Mostafa Haghir Chehreghani. Dynamical algorithms for data mining and machine learning over dynamic graphs. *WIREs Data Mining Knowl. Discov.*, 11(2), 2021.

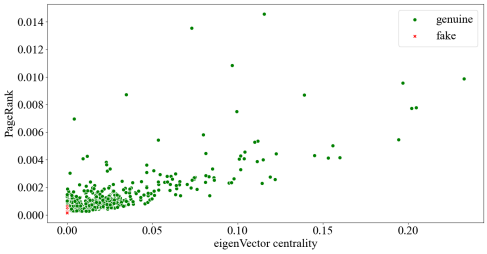
[8] Mostafa Haghir Chehreghani, Albert Bifet, and Talel Abdessalem. Discriminative distance-based network indices with application to link prediction. *Comput. J.*, 61(7):998–1014, 2018.



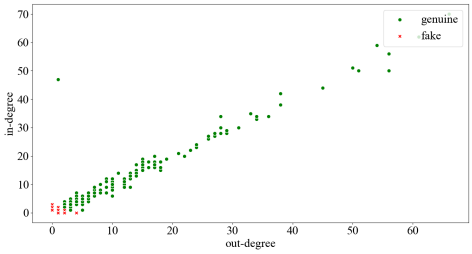
(a) correlation between *average shortest path length* and *harmonic*



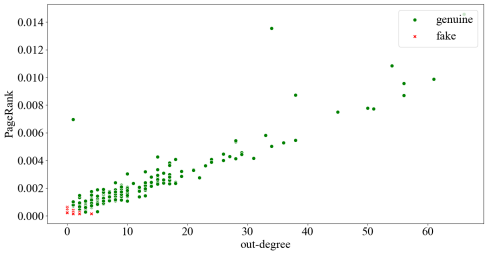
(b) correlation between *eigenvector* and *in-degree*



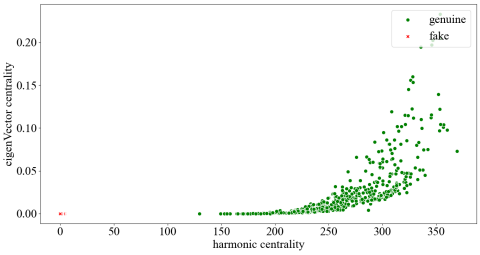
(c) correlation between *eigenvector* and *pagerank*



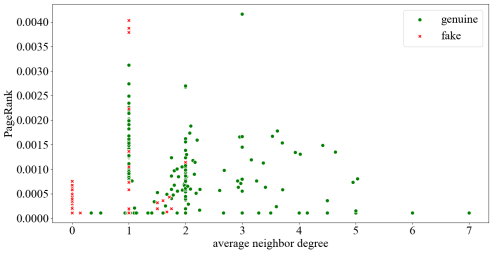
(d) correlation between *out-degree* and *in-degree*



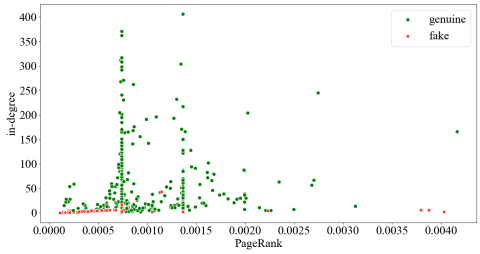
(e) correlation between *out-degree* and *pagerank*



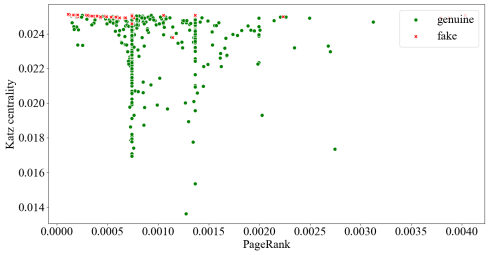
(f) correlation between *harmonic* and *eigenvector*



(g) correlation between *average neighbor degree* and *pagerank*



(h) correlation between *pagerank* and *in-degree*



(i) correlation between *pagerank* and *Katz*

Fig. 3: Green and red dots are respectively for real and fake accounts. Correlation plots 3a to 3f are taken from the follower-following graph and correlation plots 3g, 3h, and 3i are taken from the comments graph. Outliers are deleted.

[9] Mostafa Haghir Chehreghani, Albert Bifet, and Talel Abdessalem. Efficient exact and approximate algorithms for computing betweenness centrality in directed graphs. In *Advances in Knowledge Discovery and Data Mining - 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III*, pages 752–764, 2018.

[10] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. Fame for sale: Efficient detection of fake twitter followers. *Decision Support Systems*, 80:56–71, 2015.

[11] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion*, pages 963–972, 2017.

[12] Ahmad Homsi, Joyce Al Nemri, Nisma Naimat, Hamzeh Abdul Kareem, Mustafa Al-Fayoumi, and Mohammad Abu Snober. Detecting twitter fake accounts using machine learning and data reduction techniques. 2021.

[13] Sneha Kudugunta and Emilio Ferrara. Deep neural networks for bot detection. *Information Sciences*, 467:312–322, 2018.

[14] Ngoc C. Le, Manh-Tuan Dao, Hoang-Linh Nguyen, Tuyet-Nhi Nguyen, and Hue Vu. An application of random walk on fake account detection problem: A hybrid approach. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–6, 2020.

[15] Enys Mones, Lilla Vicsek, and Tamás Vicsek. Hierarchy measure for complex networks. *PLOS ONE*, 7(3):1–10, 03 2012.

[16] Arzucan Özgür, Thuy Vu, Günes Erkan, and Dragomir R. Radev. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. In *Proceedings 16th International Conference on Intelligent Systems for Molecular Biology (ISMB), Toronto, Canada, July 19-23, 2008*, pages 277–285, 2008.

[17] Sergio Porta, Vito Latora, and Emanuele Strano. Networks in urban design. six years of research in multiple centrality assessment. In Ernesto Estrada, Maria Fox, Desmond J. Higham, and Gian-Luca Oppo, editors, *Network Science - Complexity in Nature and Technology*, pages 107–129. Springer, 2010.

[18] Mohmmad Ali Abbasi Reza Zafarani and Huan Liu. *Social media mining: an introduction*. Cambridge University Press, 2014.

[19] Akrati Saxena, Pratishtha Saxena, Harita Reddy, and Ralucca Gera. A survey on studying the social networks of students. *CoRR*, abs/1909.05079, 2019.

[20] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM 2017, Montréal, Québec, Canada, May 15-18, 2017*, pages 280–289. AAAI Press, 2017.