International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (Apr-2021)
http://dx.doi.org/10.12785/ijcds/100134

Structural-Spectral-Based Approach for Anomaly Detection in Social Networks

Basim Mahmood[1] **and Mafaz Alanezi**[1]

[1] *Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq*

Received 28 Mar. 2020, Revised 9 May. 2020, Accepted 3 Aug. 2020, Published 01 Apr. 2021

Abstract: Online social networks have become one of the most effective ways for connecting and communicating with people. These networks play a significant role in our business, social, and daily activities. In these networks, people follow a particular behavior that is not necessarily identical to the actual behavior in their real life. Our goal in this work is to investigate and explore the interactions among people in the Facebook network as our targeted social network. Our investigation aims to detect the potential anomalous behaviors within the interactions among people. To this end, we involve the structural and spectral features of the network in proposing a new approach for anomaly detection. Besides, our approach is supported by concepts that are inspired from sociological theories. The data of this article was extracted from the Facebook network using Facebook Graph API. In the experimental results, the proposed approach reflected an efficient performance in terms of detecting potential anomalies and computational complexity compared to other approaches in the literature.

Keywords: Anomaly Detection, Complex Networks, Online Social Networks, Social Theories, Big Data Analysis, Data Mining

1. INTRODUCTION

Anomaly detection is an important analysis task that aims to detect abnormal data in a particular dataset of interest. It is considered an attractive area of research in the field of data analysis [1]. Anomaly detection refers to colorful aspects of detection such as outlier detection, novelty detection, deviation detection, exception mining, error detection, intrusion detection, or misuse detection [1] [2].

In fact, there are various causes for anomalies such as using data from different classes, the natural variation of data due to the nature of its distribution, data measurement errors that may occur during computations, data input error, and transmission error [2][3]. However, there are some potential causes for anomalies such that the behavior of an individual in a community. In the context of this work, we try to detect the potential anomalies based on the structural and spectral behavior of individuals in their communities. Moreover, anomaly detection can be performed using different techniques such as supervised, semi-supervised, and unsupervised [3]. However, the methodology that is used for anomaly detection is restricted by the nature of the selected dataset.

On the other side, Social Networks (SNs) have become one of the most important ways for connecting and communicating with people. The recent decade has witnessed a great revolution in SNs especially with the advent of the Internet and the great revolution of smartphones. Most of our everyday activities are organized and managed by these networks. In addition, SNs are currently considered as active platforms that are used for disseminating news, organizing events, distributing ads, spreading specific messages to people, broadcasting live events, etc. This kind of network plays a significant role in changing and socializing our societies through the disseminated contents. For instance, the Twitter social network has played an important role in the dissemination of information related social [4], education [5], and business purposes [6]. Similarly, Facebook network is considered as the main social network in many countries. It has about 2.8 billion users around the world according to Facebook officials. It is also used for different purposes such as social interactions, marketing, business, advertising, to mention a few. Furthermore, anomaly detection in online social networks has been widely investigated by researchers. There are many techniques proposed in the literature for detecting anomalous behavior. In online social networks, these techniques can be categorized into the following groups as follows [7]:

- *Behavioral-Based*: this kind of technique depends on the contents that are shared between two actors (e.g., individuals) such as messages, likes, and comments to detect anomalies.
- *Structural-Based*: depends on the structural features of data such that using centrality measurements to detect the normal and abnormal behavior of users.

E-mail: bmahmood@uomosul.edu.iq, mafazmhalanezi@uomosul.edu.iq

- *Spectral-Based*: detecting anomalies using spectral features within the space of a network such as Eigenvalues and Eigenvectors.

Our contributions in this work can be summarized as follows:

i. Investigate the potential abnormal patterns of user's interactions in the Facebook network. Studying these patterns may help us in coming up with facts on a community and utilize these facts in further considerations (e.g., marketing or business purposes).

ii. In detecting the potential anomalies, we propose to combine the characteristics of structural and spectral techniques in one new method. We strongly believe that this combination makes the detection process more accurate, comprehensive, easy, and efficient in implementation.

2. LITERATURE REVIEW

Researchers' interest in discovering anomalies dates back to the nineteenth century and was linked to the statistical community. Usually, an anomaly is seen as strange compared to others. One of the distinguished studies performed by Breunig et al. [3] claimed that for many scenarios, it is more rewarding to assign each object a degree of being an anomaly, this degree is the local anomaly factor. The reason for considering this as a local is that it depends on the degree of insulation of the object with respect to the surrounding environment. Their approach was effective for datasets where closest neighborhood queries were provided by index structures. Another important study performed by Akoglu et al. [8] discovered several new rules power-laws in density, weights, ranks, and Eigenvalues that appeared to control what was called "sub-graphs of the neighborhood". They showed how these rules were used for anomaly detection. They carefully chose features to *Oddball* design, to be scalable and could work without supervision (without user-defined parameters). They performed experiments on many real graphs with up to 1.6 million nodes, as Oddball discovers an unusual contract that matched the intuition.

Furthermore, studying anomaly detection in online social networks has strongly attracted research communities. As one of the good works was the study of Savage et al. [9], they surveyed the computational techniques used to detect anomalies in online social networks. They described anomalies as either static or dynamic, and their universe labeled or unlabeled. They reviewed methods for detecting these various types of anomalies. In their study, they proposed that the disclosure of anomalies in online social networks consisted of two sub-processes, selection and calculation network attributes, and the categorize notes from this attribute space. Hu et al. [10] investigated the difficulty of detecting structurally irregular nodes that communicated with several powerful communities in wide social networks. They found that the use of network embedding way with novel dimensional reduction technology was an effective tool for discovering such structural irregulars. They also showed that abnormalities nodes have important applications such as the enhancement of the effectual community detection. The authors in [11] determined society structures in social networks using community detection methods. For each community, node signature was merged with optimum assignment way for fitting original graph data with graph style data in order to detect two formal anomalies: node and edges anomalies. They also determined the distance between two graphs using the Euclidean formula and determined the node-to-node cost in an assignment problem by the Hungarian way for inferring the matching function. Li et al. [12] proposed a structural-based algorithm called *Radar* for anomaly detection. Their method was a learning framework that described the remaining attributes of rebuilding and its relationship to network information for detecting anomalies. By learning and investigating the remains of the rebuilding process, they were able to immediately mark anomalies in a global view when the attributes of anomalies were anonymous. Their experiments were applied using real-world datasets and produced better performance compared to baseline methods. In addition, the coherence between the remaining attribute and the network structure could help to detect anomalies other than undetected anomalies by one source of information. Yin et al. [13] proposed a spectral-based method that depended on graph spectral space in detecting network frauds and attacks. They involved Eigenvalues and Eigenvector components in their approach. In fact, the literature contains a vast number of approaches that can be involved in the anomaly detection area; Table 1 shows some of these approaches.

TABLE 1. STUDIES IN ANOMALY DETECTION AND THE TECHNIQUE THAT IS INVOLVED FOR EACH APPROACH.

Study	Technique
Stringhini et al. [14]	SVM (Support Vector Machines) Labeled dataset (spammers and non-spammers).
Akcora et al. [15]	Analyze data interest patterns of users in time.
Yu et al. [16]	Hierarchical Bayes model: Group Latent Anomaly Detection (GLAD) model.
Aswani et al. [17]	A hybrid artificial bee colony approach integrated with k-nearest neighbors.
Amato et al. [18]	Markov chains

3. DATASET COLLECTION

The dataset of this work was collected from the Facebook network for a period of two months. A crawler program called Facebook Crawler (*FC*) was designed for this purpose. FC program used Facebook Graph API for crawling public information from a variety of groups and pages. The users in these groups/pages shared their experience, knowledge, products, etc. as part of their social daily activities. The Facebook Graph API is a secure HTTP-Based API that permits developers to retrieve Facebook public data (posts or comments) using

authenticated HTTP calls. FC went through the groups/pages and compared the texts of the posts/comments with a predefined dictionary. The dictionary contains keywords that are most frequently used by people when they interact with each other informally (e.g., slang keywords in Arabic language). The main purpose of the dictionary was to reduce the amount of data retrieved as well as for network generation purposes. Then, FC retrieved the corresponding users' ids. After that, FC processed the data and stored it in a file. The strategy of FC for creating the dataset was when some of the words in the post/comment matched one or more of the keywords in the dictionary, it retrieved all the users' ids who interacted with that post (e.g., comment, like, or share). After retrieving the ids, FC considered each user-id as a node and created links among all the users who interacted with that post. The retrieved information was written in a file (output) that was used for generating our network. The generated dataset contained 27,835 nodes and 205,359 edges among them.

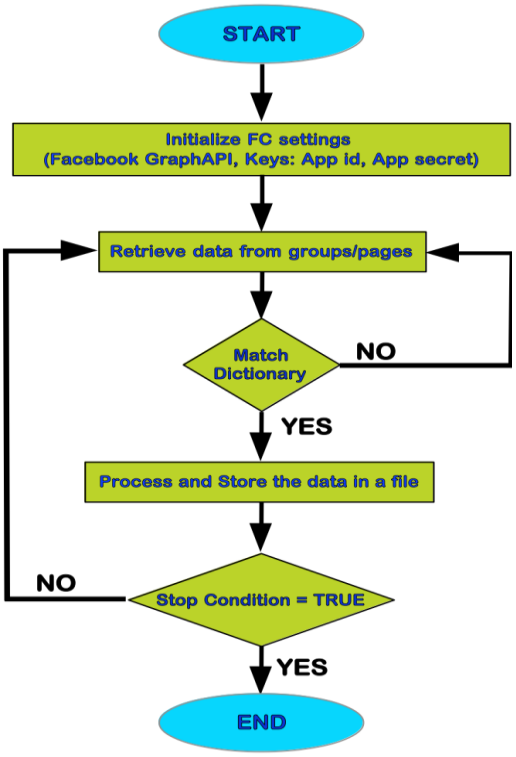

Figure 1. The flowchart of the FC crawler shows the steps of the data collection process. The process was performed using the App ID and App Secret of our developer account. For protection purposes, App ID was used to send several data requests to Facebook, while App Secret was used to decode the encrypted data.

Figure 1 depicts a flowchart of FC and how it worked.

4. NETWORK CHARACTERISTICS

The generated network was based on the collected dataset and called **I**nteractions-based **F**acebook **N**etwork (**IFN**). IFN was represented as a graph *G (V, E)*, where *V* refers to network nodes (individuals) and *E* refers to the relations among them. We extracted the characteristics of IFN using network-level measurements as follows:

- *Average Path Length l*: For all the possible pairs of individuals in a network, it is defined as the average number of paths (steps) for all the shortest paths among the pairs. In IFN, it showed the average shortest distance among individuals as follows [19]:

$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}, \qquad (1)$$

where d_{ij} is the length between the individuals *i* and *j*.

- *Diameter O*: For a network, it is the longest path among all the shortest paths [20]. In IFN, it calculated the distance between the farthest individuals.

- *Density D*: It is the proportion of the number of network edges to the number of potential (possible) edges in that network. It depicted how dense the relations among individuals in IFN and can be defined as follows [20]:

$$D_G = \frac{2(E(G))}{N(N-1)}, \qquad (2)$$

- *Average Clustering Coefficient* A_{CO}: It is also called the *global clustering coefficient*. It reflected the tendency of IFN nodes (individuals) to cluster with each other in terms of forming groups [19].

- *Communities cu*: Refers to the groups of nodes in a network that are densely connected with each other. In this work, we utilized the algorithm of Girvan-Newman [21] for clustering network nodes and extracted the number of groups/subgroups (interactive/collaborative communities) in IFN. Girvan-Newman distinguishes the edges that connect network groups, then removes these edges and keeps only the groups. This algorithm uses *betweenness* centrality measurement in the distinguishing process.

Figure 2 shows the degree distribution of IFN, it clearly followed a power-law distribution. The other characteristics of IFN can be shown in Table 2. This table also presents the characteristics of two similar networks studied in the literature.

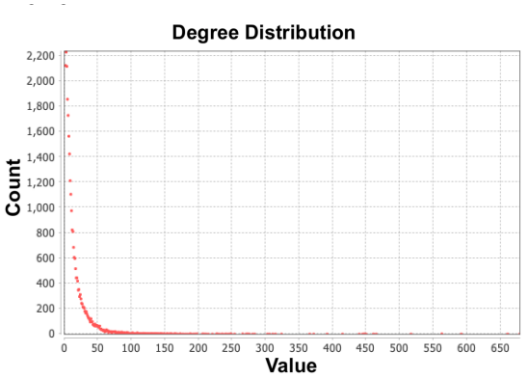

Figure2. Degree distribution of IFN network.

TABLE 2. Characteristics of IFN and two similar networks in the literature.

Network	*Nodes*	*Edges*	O	D	A_{CO}	*l*
IFN	27, 835	205, 359	15.00	0.001	0.322	4.389
SocialCircle Network [22]	144, 481	25, 696, 800	11.00	0.005	0.636	-
ONS Network [23]	48, 100	7, 840, 000	16.32	0.002	0.0471	-

The value of *O* in IFN was high compared to its number of nodes. This means that IFN contained long distances among the pairs. This is also clear compared to the average path length [24]. The value of A_{CO} tells us that the tendency of IFN users to cluster together (e.g., making groups) was not too strong, which is reasonable compared to *O*, D, and *l* values.

5. THE PROPOSED APPROACH

In developing the proposed approach, we were inspired by theories from sociology. These theories were applicable and fitted the design of our work. We involved network measurements (structural and spectral) and integrated them with sociological theories aiming at coming up with a new approach for detecting potential anomalies. We call the proposed approach **SP**ectr**A**l Structu**R**al **S**ocial-Based approach (**SPARS**).

The first step in designing SPARS was to distinguish the influential users in IFN. In this regard, we used the concept of *Elite* theory. This theory is one of the important theories in sociology. It states that a small minority of actors in a community holds the highest power in that community [25]. It reflects the fact that people with high power of relations in their society have a great impact on the majority of people in that society and they are called *Elite*. Based on this concept, we extracted the most influential users in IFN during October's demonstrations. To this end, we used a spectral measurement called (Eigenvector Centrality), which measures the importance of nodes for the connectivity of the network. In other words, it reflects how well-connected a user to the highly connected users. In fact, this measurement fitted to be incorporated in the concept of Elite theory. Therefore, our initial step was to find the elite nodes using their Eigencentralities. To perform this, given that *G (V, E)*, where *V* is a set of nodes and *E* is a set of links among them. Also, assume having an *adjacency matrix* $A = (a_{v,t})$ for the nodes *v* and *t* such that $a_{v,t} = 1$ if both nodes are connected and *0* otherwise. Then, we have *x* score for the node *v* as follows:

$$x(v) = \frac{1}{\lambda}\sum_{t\in M_{(v)}} x_t = \frac{1}{\lambda}\sum_{t\in G_{(v)}} a_{v,t}x_t \,, \qquad (3)$$

where $M(v)$ is the neighbors of node *v* and λ is the Eigenvalue. As a vector notation, the equation above can be rewritten as follows:

$$A_x = \lambda_x \qquad (4)$$

At this point, the question was: which nodes should be considered as the elite nodes in IFN? To answer this question, we investigated the values of Eigenvector and explored their distribution. Interestingly, the distribution followed a *power-law* (see Figure 3). In this kind of distribution, it is possible to apply the *Pareto* principle (or called 80/20 rule) [26]. This rule is applicable in the phenomena that are characterized by a *power-law* distribution [26]. Moreover, the Pareto rule matches very well the Elite theory [27] since it states that, "for many events, it is approximately 80% of the effects come from 20% of the causes". This means we could take the highest 20% of the Eigencentralities and consider them as the elite nodes (users or individuals) in IFN.

As mentioned, our goal in this work is to detect the potential *Anomalous Behavior Individuals* (*ABIs*) who were embedded in the groups/pages. *ABIs* are those who behave anomalously and they are relatively few individuals. Therefore, this kind of users pretended to act and behave normally [28]. Moreover, as shown in [29][30][31], *ABIs* in social networks are difficult to be distinguished and tracked. For this reason, SPARS combined spectral technique using Eigenvector centrality as well as structural technique to accurately define what was normal/ abnormal within the structure and the spectral space of IFN network.

Our next step was investigating the highest 20% of the Eigencentralities. To perform this, we decided to use the structural features of network nodes since they could deeply investigate the relations among nodes. Thus, we proposed to use other *nodes-level* measurements that could accurately contribute to detecting the potential *ABIs* in IFN. The node-level measurements enable us to deeply investigate the behavior pattern of users and also give us a view from different angle to each single user within the IFN network. These measurements are described in details as follows:

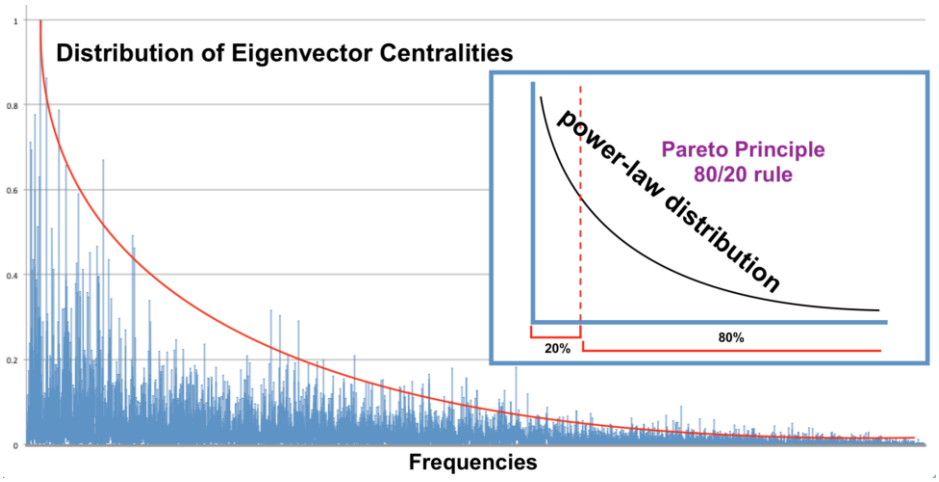

Figure 3. The distribution of Eigenvector centralities and its fitted curve (followed a power-law). The right side sub-figure describes Pareto's rule and how it was applied.

- *Clustering Coefficient* C_O: Reflects the tendency of nodes to cluster together. In IFN, *ABIs* tended to connect to particular individuals within the network and avoided connecting to others. The value of C_O can be formulated as follows [32]:

$$\mathrm{CO}_{(i)} = \frac{2|\{l_{ik}: n_j, n_k \in Ni, l_{ik} \in E\}|}{ki(ki-1)}, \quad (5)$$

where l_{jk} is a group/page between the nodes (individuals) n_j and n_k. N_i is the total IFN users and *ki* is the neighbors in the network.

- *Betweenness Centrality* C_b: Shows how many times a node appears in the shortest path of network pairs. In IFN, it reflected the importance of an individual in connecting groups or individuals. The C_b of individual *j* can be defined as follows [33]:

$$C_b(j) = \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}}, \quad (6)$$

where σ_{ik} is the shortest path between the individuals *i* and *k*. $\sigma(j)$ is the number of paths that pass through individual *j*.

- *Degree Centrality* C_d: Reflects the number of connections that a particular node has in a network. In IFN, it reflected the actual number of friends for a particular individual [32].

- *Closeness Centrality* C_c: Represents the reciprocal of the sum of all the shortest paths of a node to other network nodes. In IFN, it determined how close an individual to other individuals and can be described by [33]:

$$C_c(i) = \frac{N-1}{\Sigma_j d(ji)}, \quad (7)$$

where $d(ij)$ is the distance between the individuals *i* and *j*.

The above measurements give a deep view of the relations of a particular individual to the other individuals in IFN. Therefore, SPARS proposed to combine all these measurements in one metric called *Status* (*S*) aiming at having an indicator for detecting potential *ABIs* more precisely. In this step, we were inspired by some sociological theories such that *Collective Behavior* theory [34]. The theory describes the overall behavior of people and how they behave within their societies, groups, or communities. Based on this concept, the status *S* of an individual can be formalized as follows:

$$S(v) = (C_b(v) + C_c(v) + C_d(v))\, C_O\,(v) \quad (8)$$

where $S(v)$ is the status of individual *v* in IFN. We applied this formula to the highest 20% of the Eigencentralities.

Here, another question was raised; which nodes should be considered to be potential *ABIs*? To answer, we used the concept of *Deviance* theory [35]. This theory describes the deviant action or behavior that violates the social norms of a community or a group of people. For this reason, we considered the lowest status values as the potential *ABIs* in IFN, while the remainders were normal. Another reason, *ABIs in* IFN were not well-known users in groups and their status values could not be high under the structural-based measurements. Furthermore, detecting the lowest status values also needed more investigation. Figure 4 depicts the status distribution of the individuals, which followed a Gaussian distribution.

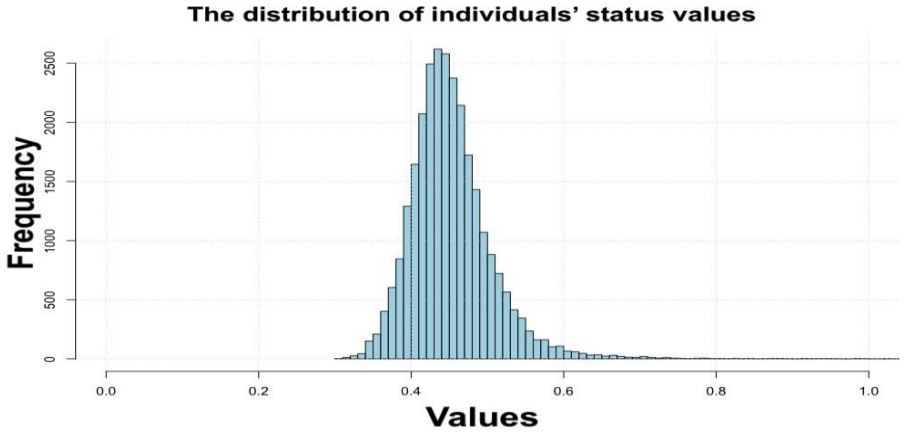

Figure 4. The distribution of the status values of the IFN nodes. The shown values in this figure represent the highest 20% of the Eigencentralities.

Based on the characteristics of Gaussian distribution, the *Empirical Rule* or called the 68 − 95 − 97.5 rule was applicable to the status distribution of individuals. This rule states that for the models that follow a Gaussian distribution there are approximately 68% of the observations are positioned between one standard deviation (σ) far from the mean (μ), 95% positioned between two standard deviations, and 99.7 between three standard deviations, they formulated as follows:

$$Pr(\mu - 1\sigma \leq \mu + 1\sigma) = 0.6827$$

$$Pr(\mu - 2\sigma \leq \mu + 2\sigma) = 0.9545$$

$$Pr(\mu - 3\sigma \leq \mu + 3\sigma) = 0.9973$$

According to the above formulas and Figure 5, we could take the values that were positioned into the left region of Part III (2.5% of individuals) and considered them as the potential *ABIs* and all the other values were considered normal. The 2.5% was originally taken from the highest 20% of the Eigencentralities. This means the potential *ABIs* represented 0.5% form the total community/ network individuals. Finally, the actual *ABIs* represented a subset of the potential *ABIs* set:

$$actualABIs \subset potentialABIs$$

To summarize SPARS, Algorithm 1 provides the steps for distinguishing the potential *ABIs* in IFN. After defining the steps of SPARS, we were able to apply it to IFN communities and evaluated the performance for each community.

6. Results and Discussion

The experimental results of SPARS were benchmarked with similar approaches in the literature using our dataset. These approaches can be summarized as follows:

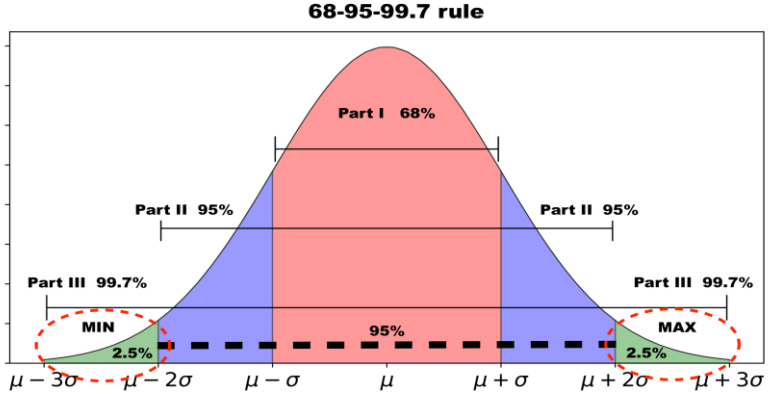

Figure 5. This figure describes the 68 − 95 − 99.7 rule of Gaussian Distribution. It illustrates how we involved this rule in IFN. As seen, the area is divided into three parts based on their distance from μ. The figure also shows the farthest distances from the mean (two-tailed regions Part III) that are both marked in red circles. Each one represents 2.5% of the total area. The left side of Part III was considered as our indicator in detecting potential anomalies.

Algorithm 1 Steps of SPARS in detecting potential anomalies in IFN.

1: **BEGIN**

2: **Foreach** $v \in G$

3: **Calculate** x_v

4: **Assign** *Elitevector* ← Highest *20%* of x $v \in G$

5: **Endfor**

6: **Foreach** $v \in$ *Elitevector*

7: **Calculate** $C_b(v)$, $C_c(v)$, $C_d(v)$, $CO(v)$

8: **Calculate** $S(v)$ using Equation 8

9: **Endfor**

10: **Assign** Potential Anomalies ← Lowest *2.5%* of $S(v) \in$ *Elitevector*

11: **END**

- Structural-Based approach: Li et al. [12] were selected as a benchmarking approach. It is also called *RADAR* and uses node centrality measurements in detecting anomalies.
- Spectral-Based approach: The work of Ying et al. [13], which is called SPECTRA, was chosen in the benchmarking since it depends on Eigenvalues and Eigenvector components to detect anomalous behaviors.
- Structural-Spectrum-Based approach: The Oddball algorithm, which is the work of Akoglu et al. [8], was important to be selected in the benchmarking. It depends on degrees, weights, and Eigenvalues in detecting anomalies.

We applied the proposed and the benchmarking approaches to the IFN detected communities and then measure the performance of each method. The community detection algorithm we used in IFN was the Girvan-Newman algorithm [21]. The main idea behind this algorithm is to distinguish the edges that hold the highest betweenness centrality values (the edges that play as bridges among communities), and then remove these edges (bridges) leaving the communities themselves. In IFN, Girvan's algorithm detected *19* main communities with a modularity level of *0.802*. In the experiments, we involved the largest five communities in IFN (*Community1* (*C*_1), *Community2* (*C*_2), *Community3* (*C*_3), *Community4* (*C*_4), and *Community 5* (*C*_5).

Figure 6 shows the performance of the approaches using our dataset. According to the obtained results, SPARS was always able to detect 0.5% of the potential anomalies in a community, considering the fact that there is always potential anomalous behavior in online social networks. *RADAR* detected about *0.1−0.35*%, *SPECTRA* detected *0.05−0.2*%, and Oddball underperformed the other approaches with *0.001−0.02*% of potential anomalies (due to its restrictive behavior in distinguishing anomalies).

As we can see, SPARS outperformed the other approaches in detecting potential anomalies that existed in the communities. Also, SPARS was simple to implement and did not need complex computations. Table 3 summarizes the performance of the approaches.

Moreover, we performed additional analysis to confirm the obtained results. In our case, we used a one-tailed t-test with 97% of confidence level ($\alpha = 0.03$). This test showed that our results were significant at the mentioned significance level compared to the obtained p-value. It is important to mention that according to our results, some of the potential anomalies were users that are originally belong to other far communities. This was based on their geodata information. This phenomenon appeared in all the five communities considered in our analysis. This means *ABIs* could not only limited to individuals who belong to their local communities but also from remote communities, which is expected since we are dealing with online networks.

TABLE 3. SUMMARIZING THE PERFORMANCE OF THE APPROACHES ON EACH COMMUNITY. THE NUMBERS SHOWN IN THE TABLE REPRESENT THE PERCENTAGES (%) OF ABIS TO THE TOTAL SIZE OF EACH COMMUNITY IN IFN.

Approaches	C*1*	C*2*	C*3*	C*4*	C*5*
SPARS	0.5	0.5	0.5	0.5	0.5
RADAR	0.35	0.25	0.3	0.18	0.1
SPECTRA	0.2	0.1	0.095	0.15	0.05
Oddball	0.022	0.019	0.015	0.02	0.001

Finally, anomaly detection in online social networks is not an easy task to perform. We believe there is no solid evidence to judge a node to be an anomaly in this kind of network. That is because of the difficulties in providing detailed information about nodes within the giant component (e.g., tracking information) as well as the restrictions in the data collection process. Moreover, most of the methods in the literature (including our benchmarking) do not provide strong proof when detecting anomalies in online social networks. Therefore, the detection process we performed in this work was on the *potential* anomalies that might exist in a community/network. This point is important to be understood since it relates to the core concept of this kind of work.

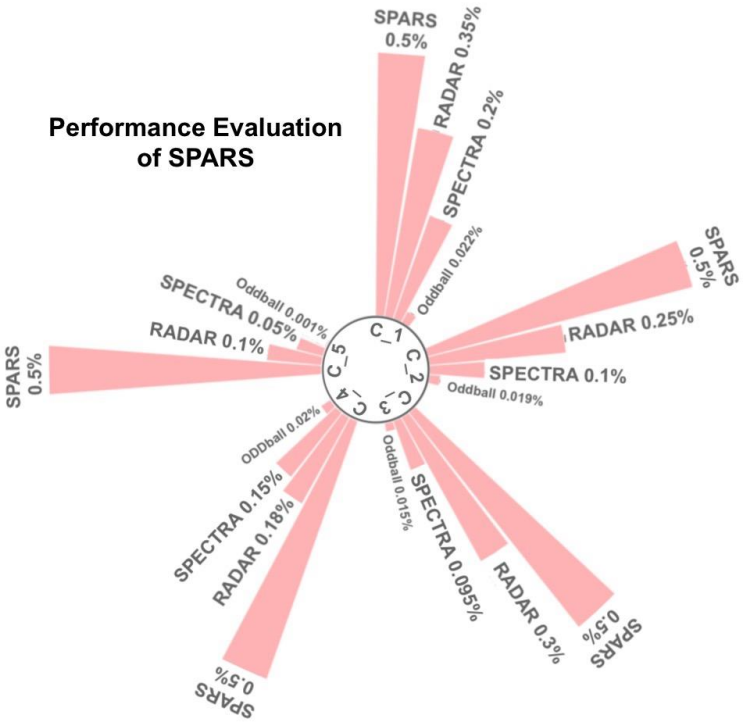

Figure 1. A circular barplot shows the performance of SPARS compared to the benchmarking approaches. The evaluation was performed for the biggest five communities detected by the Girvan-Newman algorithm. Each group of bars reflects a community in ascending order of performance.

7. Conclusions and Future Works

In this work, we developed a new structural-spectral-based approach called SPARS for detecting potential anomalies in online social networks. We generated a network for the Facebook users' interactions called IFN. The dataset used in this work was collected using a special-purpose-crawler designed for collecting data from the Facebook network using Facebook Graph API for a period of two months. SPARS used node-level measurements (Eigenvector and other centrality measurements) for distinguishing potential anomalies. We applied SPARS and the benchmarking approaches on the largest five communities in the IFN network. The experimental results showed that SPARS was able to detect 0.5% of a community as potential anomalies. In fact, the actual anomalies represented a subset (or all) of what was detected as potential anomalies. Also, the results showed that the potential anomalies in IFN communities might come from users that belong to remote communities in addition to the locals.

This work can be summarized by the following:

- The actual *ABIs* in IFN could be a maximum of 0.5% of the whole community.
- The potential *ABIs* in IFN contained users from local and remote communities.
- Investigating the structural and spectral space of a network was effective and can be simple to implement.
- Individuals in IFN reflected a weak tendency to cluster together in groups.

As future work, we plan to use IFN for testing more structural and spectral features in detecting anomalies (or potential). We also plan to investigate the correlation between the *privacy* issue and the anomaly issue among groups' members.

Acknowledgment

The researchers would like to thank the Department of Computer Science, College of Computer Science and Mathematics, University of Mosul for all the support.

References

[1] S. Pandhre, M. Gupta, and V. N. Balasubramanian, "Community-based outlier detection for edge-attributed graphs," *arXiv preprint arXiv:1612.09435*, 2016.

[2] E. Schubert, "Anomaly detection techniques: causes and issues", *International Journal of Engineering & Technology*, 7 (3.24) (2018) 449-453.

[3] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[4] Grosseck, Gabriela, and Carmen Holotescu. "Can we use Twitter for educational activities." *4th international scientific conference, eLearning and software for education, Bucharest, Romania*. 2008.

[5] Jackson, Jennifer, Sheryl Gettings, and Alison Metcalfe. ""The power of Twitter": Using social media at a conference with nursing students." (2018): 188-191.

[6] Dong, Chuqing, and Hyejoon Rim. "Exploring nonprofit-business partnerships on Twitter from a network perspective." *Public relations review* 45.1 (2019): 104-118.

[7] Centola, Damon. *How behavior spreads: The science of complex contagions*. Vol. 3. Princeton, NJ: Princeton University Press, 2018.

[8] L. Akoglu, M. McGlohon, and C. Faloutsos, "Oddball: Spotting anomalies in weighted graphs," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2010, pp. 410–421.

[9] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," *Social Networks*, vol. 39, pp. 62–70, 2014.

[10] R. Hu, C. C. Aggarwal, S. Ma, and J. Huai, "An embedding approach to anomaly detection," in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*. IEEE, 2016, pp. 385–396.

[11] S. Mekouar, N. Zrira, and E. H. Bouyakhf, "Community outlier detection in social networks based on graph matching," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 11, no. 3, pp. 209–231, 2018.

[12] J. Li, H. Dani, X. Hu, and H. Liu, "Radar: Residual analysis for anomaly detection in attributed networks," in *IJCAI*, 2017, pp. 2152–2158.

[13] X. Ying, X. Wu, and D. Barbar´a, "Spectrum based fraud detection in social networks," in *2011 IEEE 27th International Conference on Data Engineering*. IEEE, 2011, pp. 912–923.

[14] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proceedings of the 26th annual computer security applications conference*, 2010, pp. 1–9.

[15] C. G. Akcora, B. Carminati, E. Ferrari, and M. Kantarcioglu, "Detecting anomalies in social network data consumption," *Social Network Analysis and Mining*, vol. 4, no. 1, p. 231, 2014.

[16] R. Yu, X. He, and Y. Liu, "Glad: group anomaly detection in social media analysis," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 2, pp. 1–22, 2015.

[17] R. Aswani, S. Ghrera, A. K. Kar, and S. Chandra, "Identifying buzz in social media: a hybrid approach using artificial bee colony and k-nearest neighbors for outlier detection," *Social Network Analysis and Mining*, vol. 7, no. 1, p. 38, 2017.

[18] F. Amato, A. Castiglione, A. De Santo, V. Moscato, A. Picariello, F. Persia, and G. Sperl´ı, "Recognizing human behaviours in online social networks," *Computers & Security*, vol. 74, pp. 355–370, 2018.

[19] D. J. Watts and S. H. Strogatz, "Collective dynamics of âĂŸsmall-worldâĂŹnetworks," *nature*, vol. 393, no. 6684, p. 440, 1998.

[20] R. Albert and A.-L. Barab´asi, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[21] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[22] A. De Salve, M. Dondio, B. Guidi, and L. Ricci, "The impact of userâĂŹs availability on on-line ego networks: a facebook analysis," *Computer Communications*, vol. 73, pp. 211–218, 2016.

[23] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Crawling facebook for social network analysis purposes," in *Proceedings of the international conference on web intelligence, mining and semantics*, 2011, pp. 1–8.

[24] G. Scardoni and C. Laudanna, "Centralities based analysis of complex networks," *New Frontiers in Graph Theory*, pp. 323–348, 2012.

[25] T. Bottomore, *Elites and society*. Routledge, 2006.

[26] M. E. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[27] R. A. Nye, *The anti-democratic sources of elite theory: Pareto, Mosca, Michels*. Sage Publications, 1977, no. 21.

[28] S. Anna, "The information warfare aspect of the syrian conflict in the context of the us-russia power struggle in the middle east," 2018.

[29] J. Moskvina *et al.*, "Their fight, our news: A case study of pussy riot in british and russian online media," *Medi´aln´ı studia*, vol. 13, no. 2, pp. 123–142, 2019.

[30] J. DeNardo, *Power in numbers: The political strategy of protest and rebellion*. Princeton University Press, 2014.

[31] H. Duschinski, "Reproducing regimes of impunity: Fake encounters and the informalization of everyday violence in kashmir valley," *Cultural Studies*, vol. 24, no. 1, pp. 110–132, 2010.

[32] R. Menezes, A. Evsukoff, and M. C. Gonz´alez, *Complex Networks*. Springer, 2013.

[33] B. Mahmood and R. Menezes, "United states congress relations according to liberal and conservative newspapers," in *Network Science Workshop (NSW), 2013 IEEE 2nd*. IEEE, 2013, pp. 98–101.

[34] N. J. Smelser, *Theory of collective behavior*. Quid Pro Books, 2011.

[35] R. L. Matsueda, "Cultural deviance theory': The remarkable persistence of a flawed term," *Theoretical Criminology*, vol. 1, no. 4, pp. 429–452, 1997.

Basim Mahmood He is a faculty member at the Department of Computer Science, University of Mosul, IRAQ. He obtained his Ph.D. degree in Computer Science in the field of Complex Networks from Florida Institute of Technology/ USA in 2015. His M.Sc. degree was also in Computer Science in the field of Mobile Systems from the University of Mosul/ IRAQ in 2009. His current area of research deals with Complex Networks applications, Big Data Analysis, and Internet of Things (IoT).

Mafaz Alanezi She is a faculty member at the Department of Computer Science, University of Mosul, IRAQ. She obtained her Ph.D. degree in Computer Science in the field of Computer and Network Security from University of Mosul / Iraq in 2012. Her M.Sc. degree was also in Computer Science in the field of Image Processing from the University of Mosul/ IRAQ in 2003. Her current area of research deals with Computer and Network Security, Artificial Intelligence, and cloud computing.