Computing and Informatics, Vol. 41, 2022, 1358–1382, doi: 10.31577/cai_2022_5_1358

PRIVACY ISSUE: FROM STATIC TO DYNAMIC ONLINE SOCIAL NETWORKS

Mafaz Alanezi, Basim Mahmood

Computer Science Department
University of Mosul
Mosul 41002, Iraq
e-mail: {mafazmhalanezi, bmahmood}@uomosul.edu.iq

Abstract. Today's societies have become more dependent on social networks in terms of communications and interactions. These networks contain most of the people's activities, which can be public or even personal events. In the last decade, social networks have turned into more prominent platforms in managing and organizing public events. The Egyptian revolution in 2011 and the Ukrainian revolution in 2014 are good reflections of such events. However, it is not known how much the privacy issue of users is revealed in the reality as a consequence of their online interactions. In this work, we investigate the privacy issue in online social networks and its reflection on real life. Our dataset was extracted from the Facebook groups/pages that were involved in the 2019 Iraqi October revolution. Our approach generates a static network using the collected dataset. Then, we investigate the generated static network in terms of detecting potential anomalies. After that, we project the static network (including its characteristics) into a dynamic environment and generate a dynamic network for investigating the privacy issue in the real life. The contribution of this work lies in projecting a real-world static network into a dynamic environment aiming at investigating users' privacy in the real world. Finally, this kind of approach has not been given enough attention in the literature and it is therefore deeply investigated in this article.

Keywords: Privacy issue, online networks, projecting static to dynamic networks

1 INTRODUCTION

Online social networks have become a fundamental need in modern life [1]. Today, many of our activities are managed and organized using these networks such as Twitter, Facebook, WhatsApp, Instagram, etc. The number of users in these networks varies from a country to country. For instance, some countries' users prefer using Facebook more frequently, while other countries' users prefer to use other online social networks. The main characteristics of these networks are the ability of users to share their personal content with friends, show their opinions, interact with others, and make comments [2]. Also, these networks enable users to configure their accounts based on their preferences [3]. However, social networks still struggle with the issue of privacy because users are exposed to each others' accounts and it is difficult to maintain users' privacy. Although these networks do not share users' information with the public, some online applications request users' permission for accessing their information and further use it for marketing or advertising purposes [4]. The presence of users in online social networks leads to a trade-off between the possibility of enlarging their social and professional circles, and privacy threats. Moreover, in online social networks, users may provide information (sensitive/insensitive) to third parties; this information may be revealed to other users and violate the privacy. The information may include spatial or temporal elements such as their location and time stamp, or personal characteristics such as personal background, hobbies, contacts, personal views, etc. Consequently, this sharing habit can be a reason of potential threats for users (e.g., identity theft, sexual assault, stalking, hiring, online abuse, surveillance, and unintended fame and even deceptive ads) [5]. On the other hand, online social networks have broken the real-life spatiotemporal barriers due to the low communication cost. This phenomenon has made online social networks to be one of the influential factors in real life. This means, our real-life activities are directly influenced by our interactions in the social networks.

The analysis of online social networks depends on the structural nature of these networks, which can be either static or dynamic [6, 7]. The traditional kind of analysis has concentrated on the representation of graphs as static networks. In this case, it is more likely that researchers tend to use community detection approaches or identifying worthwhile group structures in the network. In contrast, dynamic networks' nodes are not stationary and their positions change over time. This means each node is subject to spatial and temporal aspects in its features. Furthermore, our real-life environment is considered a dynamic network in which individuals represent the nodes and the connections among them are driven by their social relations. Practically, when simulating a dynamic network, it is needed to incorporate a particular mobility model that reflects the movement patterns of the mobile nodes. Therefore, the analysis of dynamic networks is more challenging than static networks [8]. Besides, in dynamic networks, the pattern of reactions is also changed over time, for instance "who interacts with whom" at time t. In the same context, another kind of network is called Labeled networks, which include informa-

tion on different characteristics of individuals and their interactions [7]. The analysis of these networks needs to model the structural changes in parallel with the change in time.

Investigating dynamic networks has been given enough attention in the literature. Several approaches have been proposed for detecting a variety of phenomena (e.g., anomalies) in these networks [9]. The study of Liao et al. [10] proposed a tool for investigating anomalies in dynamic networks. The tool uses the spatial and temporal dimensions of network nodes to analyse anomalies. Another study performed by Chen et al. [11] used a recurrent neural network in proposing a new differential privacy scheme. The authors investigated the privacy issue for users in a real-time manner. They showed that the proposed approach has the ability to protect the privacy of user's data. However, their approach was limited since it analysed a network as snapshots at a specific time. Kokciyan and Yolum [6] classified the privacy violations that occur in online social networks. They found that privacy violations in these networks arise from complex interactions. To understand these violations it was needed for a semantic understanding of the occurred events. Therefore, they suggested an agent-based representation of a social network, where agents manage user privacy requirements by generating obligations with the system. The suggested detection algorithm performed the analysis using the logical description at various depths of realistic social networks. The main limitation of their work was the size of data that can be processed and analysed. Cho et al. [5] studied the ability of users in enlarging their social networks by making reliable friendships while not infiltrate their private information to unofficial individuals or social attackers. They applied the notions of confidence and reputation for maintaining the privacy of users while improving their social capital in online social networks. Using the social network topology from Facebook, they designed a template for individual user interactions with other users which depends on feeding such as posting information and behavior of reactions (e.g., providing likes or comments). Their outcomes show that there is a trade-off between social capital and maintaining privacy. The study struggled with the huge amount of data that are generated as well as the accuracy of users' history of interactions. Bhagat et al. [12, 13] studied the problem of frequent deployment of online social networks data as network growth while maintaining users' privacy. They used link prediction algorithms to model the development. They also proposed to mask a dynamic network identity when new nodes and edges are added to the network. The prediction graph was used for group-based anonymity. Their approach contributed to protecting the privacy and predicting network evolution but it was difficult to predict future privacy violations.

Another study by Kafali et al. [3] developed an approach called PROTOSS as a run-time tool for detecting and predicting privacy violations in online social networks. Their approach detected the relationships among users, their privacy agreements with the network operator, and domain-based semantic information and rules. PROTOSS was used to discover whether relationships among users will violate privacy agreements. The approach was also able to expect potential future viola-

tions through feeding in a hypothetical future global situation. By experimenting the model on the scenario as well as on the existing Facebook dataset, PROTOSS could detect and predict exact leaks similar to those reported in real-life examples. The main limitations of PROTOSS were the lack when dealing with a large scale of data and the realistic of relations among the used concepts. Furthermore, static and dynamic networks can be used for analyzing the privacy issue among people. However, when it comes to investigating privacy based on the interactions among people, it is better to use a dynamic network, as discussed in the study of Farine [14]. He studied static and dynamic networks and proved that dynamic networks are better for predicting the behavior of people based on their temporal interactions.

According to the literature, most of the approaches have limitations when predicting future potential privacy violations. More precisely, these limitations are circled around the difficulties when dealing with a large scale of data (e.g., users and their interactions) due to the heavy simulations that, in turn, generates big data. The other limitation is that most of the researchers investigate the privacy issue either in a static or in a dynamic network. In this context, there is a severe lack in developing approaches that consider the history of interactions among people in their static online networks and further utilize it in their dynamic network when investigating the privacy issue. Investigating the privacy issue in real life needs a lot of attention and many factors should be considered during the investigations. Furthermore, we believe that the interactions among people in online social networks are not enough to be considered as the main factor when it comes to privacy issue. Other factors such as people interactions in their real environment can also be considered. Therefore, it is important to integrate the characteristics of people interactions in their both online social networks and the dynamic environments they are practice their life in. Hence, to deal with the aforementioned gaps, our contributions are:

- Develop a novel framework for projecting a real online static network into a dynamic environment for tracking purposes.
- Use a large-scale dataset that includes users and their features (e.g., relations and interactions) to investigate future potential privacy violations of users considering their history of interactions in static networks, and their interactions in the dynamic environment.

The strength of our approach is the ability to project any large-scale online social network into a dynamic environment. Also, it can predict the future potential privacy violations of users as a consequence of their interactions in the online social networks. The applications of our approach can be in investigating a wide range of aspects such as privacy violations, anomaly detection, tracking, and health-related applications.

This article is organized as follows: the next section includes our research method and how this work was performed step by step. In Section 3, we present the obtained

results and discuss them in details. Finally, we conclude our work in Section 4 and present the main challenges and the future works.

2 RESEARCH METHOD

2.1 Dataset Collection

Our dataset was collected during the period of the recent Iraqi demonstrations (October 15, 2019 to December 15, 2019). The dataset was extracted from Facebook groups/pages that were mainly involved in organizing and managing the demonstrations' events. The data collection process was performed using a special-purpose crawler, which was designed for this work. The collected texts (posts/comments) were compared word-by-word to a pre-designed dictionary (see Figure 1). This dictionary contained the most frequent words used by the Iraqi people (Iraqi slang and informal words) during the demonstrations. The collection process was based on a particular strategy; if the contents of the selected groups/pages match words in the dictionary, it means the post was mainly related to the demonstrations. In this case, we collected users' information such as user-id, location, and other information. All the collected information was related to the users who interacted with the collected posts and then connected them. In other words, we represent users as nodes, if two users (nodes) have interacted (e.g., comment) with the same post; this means they have an edge between them. This strategy was applied to all nodes in the generated network. It should be mentioned that we use Facebook Graph API in the collection process. This API is a secure HTTP-based API that enables retrieving Facebook public data using authenticated HTTP calls. After that, we processed the structure of the collected data to be suitable for visualization and analysis. The number of nodes in our dataset was 27 835 representing the Facebook users who interacted with the demonstrations and 205 359 edges connecting them. Now, we have a network with nodes and edges that reflect the online relations and interactions among Facebook users. In fact, the actual number of users (demonstrators) was significantly greater than the collected users, but we believe this large scale of data is sufficient enough for our study.

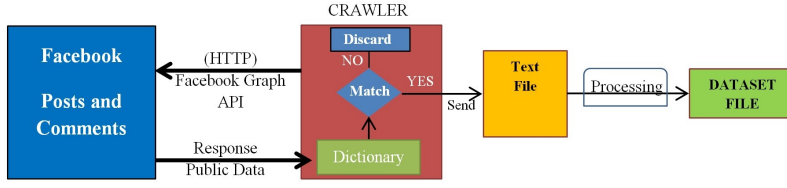


Figure 1. Data collection process

2.2 The Proposed Framework

In this work, we propose a new framework to investigate future potential privacy violations based on people's relations and interactions. To this end, we first generate a static network using the collected dataset and detect potential anomalies in people behavior. Then, we project each node in the static network (including its characteristics) into a dynamic environment aiming at generating a dynamic network. Then, we use the dynamic network for investigating the privacy issue based on 1) the interactions of people within the simulation environment, 2) the information that was inherited from the static network (history of interactions). In fact, violating the privacy of a user can be performed in different ways. For instance, how frequently a user interacts/communicates with anomalous users. This kind of interaction can be a significant factor for violating the privacy of users. Also, for a user, a high level of interactions with others may lead the privacy to be more exposed to others. Therefore, in this work, we take into considerations the history of interactions in the static network and the interactions of users in the dynamic environment for evaluating the privacy. We believe this is a good practice because when simulating a dynamic environment, it is better to have the history of interactions of people when perform the simulations, which is reasonable for this kind of analyses. Moreover, our approach has the ability to assess the impact of users' interactions in their social networks on their privacy in real life. Practically, the proposed approach has three main steps, as we explain in the following subsections.

2.2.1 Step 1: Detect Potential Anomalies in the Static Network

In this step, we are inspired by sociological concepts and theories. Structural and spectral measurements are involved and integrated with the theories aiming at detecting potential anomalies. Now, we start by distinguishing the influential users in our network. To do this, the concept of Elite theory is involved. Elite theory is considered one of the powerful theories in the field of sociology. It states that "a small minority of actors in a community holds the highest power in that community" [15, 16]. It has the ability to distinguish people with high-power of relations in a community and is called Elite people. By the means of this theory, our approach distinguished the most influential users in the static network during the demonstrations. A spectral measurement (Eigenvector Centrality) is involved to evaluate the importance of a user in network connectivity. More precisely, Eigencentrality shows how well-connected a particular user is to users with high connections. This centrality measurement fits very well with the concept of Elite. Therefore, the elite users are those who have the highest values of Eigencentrality. To calculate it, given a graph $G(V, E)$, where V denotes a set of nodes and E denotes a set of edges connecting the nodes. We also assume an adjacency matrix A, where $A = (av, t)$ for the nodes v and t such that av, $t = 1$ if both nodes are connected and 0 otherwise. The x score for node v is formulated as

follows [17]:

$$x(v) = 1/\lambda \sum_{t \in M_{(v)}} x_t = 1/\lambda \sum_{t \in G} A_{(v,t)} x_t \tag{1}$$

where $M(v)$ refers to the neighbors of v and the Eigenvalue is λ. As a vector notation, the formula (1) can be reformulated as follows:

$$A_x = \lambda x. \tag{2}$$

Now, we have to distinguish the elite users in our network. Before that, it is, first, needed to investigate the distribution of Eigencentrality. We found that the distribution follows a power-law as depicted in Figure 2. One of the interesting facts of this kind of distribution is the possibility of applying the Pareto rule (80/20 rule) [18]. This principle is applicable for all the applications that are characterized by a power-law according to [18]. Interestingly, this rule fits very well with the concept of Elite theory since it states: "for many events approximately 80 % of the effects come from 20 % of the causes". Hence, our approach is able to consider the highest 20 % of the Eigencentrality of users as the elite users in our network.

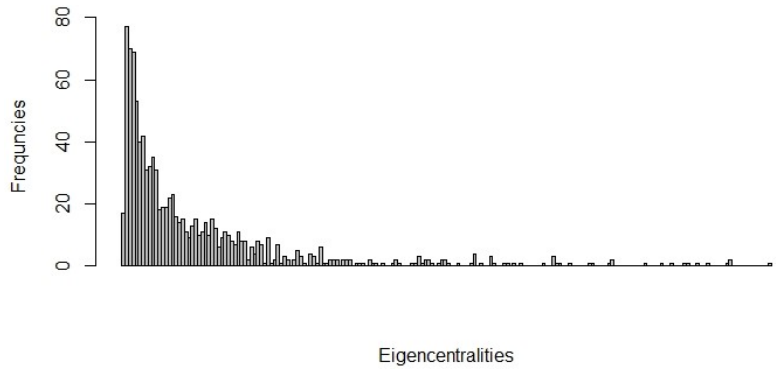


Figure 2. Distribution of eigencentralities

In this paper, we will frequently use the term ABIs to mention people with abnormal behavior. This term is the acronym of Anomalous Behavior Individuals, which are few people who had behaved abnormally/anomalously in their online interactions during the demonstrations (or any other events) [19]. Moreover, distinguishing/tracking ABIs is considered as a challenging task due to the lack of information about them. Therefore, in addition to use a spectral measurement (Eigencentrality), we also use structural measurements to more accurately define the normal and abnormal behavior within the structural and spectral spaces of our network.

According to the aforementioned, we decide to investigate the highest 20 % values of the Eigencentrality. This investigation is performed using the structural space of network users. The structural measurements can dig deeply into the relations among users. Therefore, we propose to use nodes-level measurements, which we believe they can be a powerful tool and contribute to the detection of potential ABIs as follows.

- *Clustering Coefficient* C_O: it is a reflection of the tendency of users to group with other network users. By the means of our network, ABIs tend to be connected to specific users. The value of C_O for undirected network is formulated as follows [20]:

$$C_O(i) = \frac{2|\{l_{jk} : n_j : n_k \in N, l_{jk} \in E\}|}{k_i(k_i - 1)} \tag{3}$$

  where l_{jk} is a group/page between users n_j and n_k, and N denotes the total users in the network, and k_i denotes the number of neighbors.

- *Betweenness Centrality* C_b: it reveals how many times a user appears in the shortest path between any given pairs of users within the network. In other words, it shows how well-positioned (e.g., importance) a user in connecting groups or network users. C_b of user j is defined as follows [20]:

$$C_b(j) = \sum_{i \neq j \neq k} \frac{\sigma_{ik}(j)}{\sigma_{ik}} \tag{4}$$

  where σ_{ik} denotes the shortest paths between the users i and k. $\sigma(j)$ denotes the number of network paths that pass through user j.

- *Degree Centrality* C_d: it is the frequency of connections of a user, which means the actual number of friends for that user [20].

- *Closeness Centrality* C_c: it is the user's reciprocal of the sum of all the shortest paths to other users. It reflects how close a user is to other network users and it is formulated as follows [20]:

$$C_c(i) = \frac{(N-1)}{\sum_j d(ij)} \tag{5}$$

  where $d(ij)$ denotes the distance between the pairs i and j.

These measurements enable us to provide a multidimensional view of the actual relations among network users. Our proposed approach aims to combine these measurements in what we call a user's Status (S). Our inspiration at this point comes from sociological concepts and theories (e.g., Collective Behavior Theory [21]). Based on this theory, we can explain the behavior of users within their communities. The S values, then, will be used as indicators in distinguishing potential ABIs. The status S of a user is proposed to be as follows:

$$S(v) = (C_b(v) + C_c(v) + C_d(v))^{C_O(v)} \tag{6}$$

where $S(v)$ denotes a user v's status and the other parts of the equation are already defined in the previous paragraphs. The main idea behind the above equation is to extract the actual structural behavior of users within the static network, which is crucial in determining how well-positioned the users are in the network. We plan to apply this equation to the highest 20 % of the Eigencentrality.

Now, for distinguishing the potential ABIs, the concept of Deviance theory [22] is involved. It explained the "deviant action or behavior" that violates communities' social norms. Based on this concept and the concepts of the parameters of Equation (6), the lowest S values (abnormal) will be considered as the potential ABIs. Normal users in online social networks are usually considered to be almost better-positioned because they already have their circles/groups and almost connected to people they know. On the other hand, abnormal users are usually considered intruders to the circles/groups and they are not well-positioned within the network. Now, there is a need for an indicator that helps us to decide which values are the lowest. Figure 3 shows the distribution of S values; clearly, they follow a Normal distribution. One of the important features in this distribution is the "Empirical Rule" [23], which is applicable to S values. The Empirical rule states that for all the phenomena that are characterized by a Normal distribution, it is approximately 68 % of the observations are located within a distance of one standard deviation (σ) from both sides of the mean (μ), 95 % within (2σ), and 99.7 % within (3σ). According to the aforementioned, we decide to consider the values that are located in the left region (2.5 % of users) as the potential ABIs users (see Figure 3). The mentioned percentage (2.5 %) is originally extracted from the highest 20 % of the Eigencentrality. The precise percentage of the potential ABIs users equals 0.5 % of the total network users.

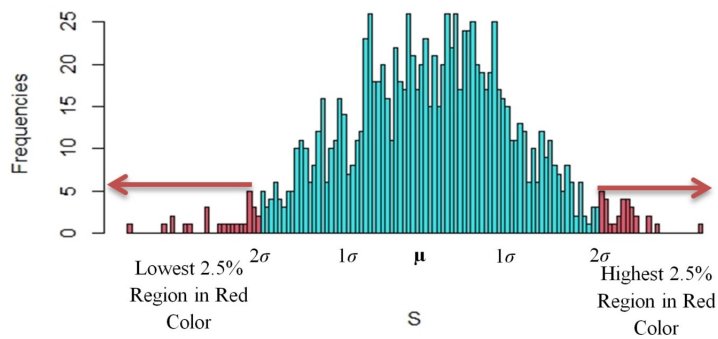


Figure 3. Distribution of S values

2.2.2 Step 2: Projecting the Static Network into a Dynamic Environment

In this step, we project the generated static network into a dynamic environment. Performing this step is not an easy task. It needs a lot of attention because simulating dynamic networks needs many requirements such as the nature of the environment that will be used in the simulation, the mobility model that should be incorporated to describe the movement patterns of the mobile objects (users), the distribution of users within the environment, and the strategy of tracking users in the environment. These requirements should be defined and prepared before performing the simulations.

Nodes in static networks are stationary and positioned in fixed locations. A particular topology can be involved to describe the positions of nodes within an environment. However, in dynamic networks, the positions of nodes are not fixed and are changed over time. Practically, it is important, for simulation purposes, to use a particular mobility model that describes the movement patterns of nodes in terms of direction and velocity. In this context, developers should use models that accurately reflect the movement patterns of the mobile nodes, which can be people, animals, automobiles, etc. In our work, the mobile nodes are humans. Therefore, in the proposed framework, we include a mobility model that accurately simulates people movement patterns. In this regard, we propose to use the model that was developed by Song et al. in 2010 [24]. His model is the most known accurate model that reflects human movement patterns and is called Individual Mobility (IM) model. It is based on two main mechanisms. *Exploration*, which states that the number of explored locations is decreased by time. More precisely, as time goes, the number of locations that a particular individual visit is decreased. The probability (P_{new}) of exploring a new location is formalized as:

$$P_{new} = \rho L^{-\lambda} \tag{7}$$

where ρ and λ are used to control the tendency of users to explore new locations in the next move (step) and L is the number of explored positions (locations). The second mechanism is *Preferential Return*, which means people tend to return to the most explored locations they have visited in the previous movements (past) with a complementary probability (P_{ret}) of the previous mechanism as follows:

$$P_{ret} = 1 - P_{new}. \tag{8}$$

It should be mentioned, in the IM model as people move, the number of explored locations L is increased by 1 ($L = L + 1$) in the next move. In this work, our dataset contains two kinds of nodes; regular users and potential ABIs. In the simulations, the former should be simulated normally in terms of their movement patterns. Therefore, we use the IM model to simulate their dynamics. The latter (ABIs) are expected (as we propose) to behave differently in terms of their movements patterns. Therefore, minor changes should be made on the IM model when simulating this kind of user. In real life, abnormal behavior can be observed in

revolutionary events. ABIs, in this kind of event, tend to explore and visit more locations than usual. They have a strong tendency to collect more information on the events by visiting as many places as they can. For this reason, we propose to perform some modifications and re-formalize the first mechanism (exploration) in Song's model. The updated mechanism should increase the number of explored locations to be close to real-life behavior. To this end, we propose to add a parameter called N^+. This parameter is used to increase the number of explored locations. Therefore, the probability of exploring new locations for the ABI ($P_{ABI-new}$) can be formalized as follows:

$$P_{ABI-new} = \rho(L + (1/N^+))^{-\lambda}. \tag{9}$$

Practically, N^+ is an integer dynamic-variable that is changed over time. It adds some randomness pattern to the whole movement pattern, which is desired since we involve ABIs. During the simulations the value of N^+ is randomly selected at time t:

$$L(t) > N^+(t) > 0. \tag{10}$$

To summarize, we use the IM model for simulating the regular users, while in simulating ABIs, we use the proposed updated version of the first mechanism in the IM model.

The second requirement for our simulations is the tracking procedure that should be used to monitor users' movements and their dynamic interactions. The process of projecting users from the static to the dynamic environment (Projection Strategy) is based on a relation-driven approach. We propose that users are distributed within the simulation environment based on their relations to each other in terms of positions. For instance, "friends" in static networks are positioned close to each other. Therefore, at time $t = 0$ in the simulations, the groups in the static network are projected in the dynamic environment and positioned close to each other. During the simulations, for instance, at time $t = 1 - n$, the movements of users are driven by the IM model. Figure 4 shows an example of how the projection is performed. As the simulations in running, users move within the environment and encounter each other; they change their positions and directions. All the information that is related to the interactions among users is reported. This means we can keep track of every single user in the dynamic network. Tracking information includes: with whom users encounter, how regular these encounters are, how long each encounter lasts for, the number of visited locations, and other interactions-related information. This kind of information will further be utilized when evaluating the privacy in the dynamic networks. The other requirement for our simulations is the way of deploying users within the simulation environment. In this regard, the above-mentioned projection strategy is not enough and it is needed to follow a particular pattern/method for deploying the users within the simulation environment. According to the distinguished work of Grossman [25], in metropolitan cities, people are deployed based on a power-law distribution. This means, people are concentrated in the city center and gradually decreased where approaching city borders. Therefore, in our simulations, all users are deployed based on

a power-law distribution with respect to the projection strategy in the previous paragraph.

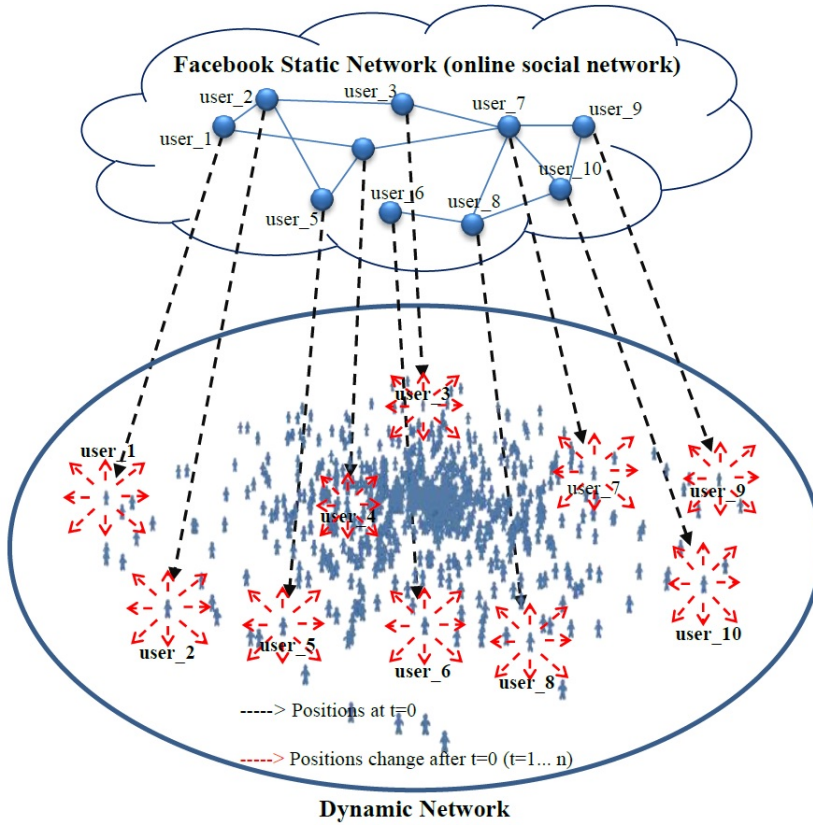


Figure 4. An example of our Projection Strategy and how the static network is projected into a dynamic environment to generate a dynamic network

2.2.3 Step 3: Analyse the Dynamic Network in Terms of Privacy

In this step, we aim to analyse the interactions among people based on the generated dynamic network. To this end, it is needed to describe the characteristics of human relations and interactions. Investigating the relations among people is important insofar as it contributes to measuring the privacy issue, which is our purpose in this work. According to the distinguished work of Barabasi [26], people relations and interactions can be characterized by several features. In this work, our proposed approach is based on three features that mainly characterize the relationship between two users. Now, the first feature is the Recurrence of Meetings (RM), which reflects how many two users meet. The second feature is the Consistency of Meetings (CM),

which means how consistent/regular the meetings between two users are. It can be calculated using the average time between every two meetings for a pair of users as follows:

$$CM_{ij}(t) = \sum_{t=1}^{t_{current}} \frac{wait_{ij}(t)}{RM_{ij}(t_c)} \tag{11}$$

where CM_{ij} denotes the consistency of the meetings for the users i and j at time t, and $wait_{ij}$ denotes the waiting time of all the meetings between i and j at time t. $t_{current}$ represents the time of the current meeting and RM_{ij} is the recurrence of meetings. The third feature we propose to use is the Duration of Meetings (DM), which is the average of the duration times of all the meetings between two users and can be formulated as follows:

$$DM_{ij}(t) = \sum_{t=t_f}^{t_{current}} \frac{du_{ij}(t)}{RM_{ij}(t)} \tag{12}$$

where $DM_{ij}(t)$ denotes the average duration time of all the meetings between the users i and j at time t. $du_{ij}(t)$ denotes the duration of the meeting between i and j at time t, t_f denotes the time of the first meeting between i and j. Based on the aforementioned features, we calculated the weight of the relations among every single pair of users in our network. The weight (W) of the relation between the users i and j represent the collected values of the three aforementioned features as follows:

$$W_{ij}(t) = RM_{ij}(t) + DM_{ij}(t) + (1/CM_{ij}(t)). \tag{13}$$

The term $(1/CM_{ij}(t))$ makes it fair enough for the three features when calculating the weights. More precisely, if two users meet every long period of time, we cannot say they have a strong relation. Therefore, a low value of CM reflects a strong relation between two users. It should be mentioned that the weight of the relation between two users is dynamic and recalculated whenever an interaction happens. Also, all the three mentioned features are normalized to be between 0 and 1.

3 RESULTS AND DISCUSSIONS

For the dynamic network, we performed our simulations using a special-purpose simulator that is used for this kind of work. The simulator is called *Social-Network-of-Sensors* (*SNoS*) that is based on the multi-agent programming of NetLogo modeling. The kind of simulations performed in this work is similar to the simulations performed in [27]. The simulation environment is designed as a city with borders and partitioned into blocks (patches). The projected users are deployed based on power-law distribution. The movement of users is driven by the IM model. During the movements of users within the environment, they cannot pass city borders because the steps in the IM model are dominated by exponential-cut-off. For the sake of accuracy, we carried out the simulations 20 times then we

averaged them. In the simulator, each step is called a *tick* that is approximately equal to 1.2 minutes in reality considering the human walk speed of 3 km/h according to Grossman [25]. We proposed to simulate our approach for the same period of the dataset collection, which was for 67 200 ticks simulating 80 640 minutes (8 weeks) in the reality. The reason behind selecting this period is to show the impact of both networks under the same time constraints. The distance between every two users is recalculated at every step. In our simulations, two users meet if they stop on the same block (location) and leave that block at the same time.

The first step towards our results was the evaluation of the privacy issue in the static network. The evaluation was based on the frequency of interactions among users. In other words, how many times two users had interacted with the same post (e.g., comment, like, or share). The evaluation was performed among all the pairs of regular (normal) users as well as between the regular users and the ABIs in the network. This means, we evaluate the level of interactions among the regular-regular pairs and compared it regular-ABIs pairs. The main purpose of this evaluation was to reveal the behavior of the static network and to further use it as an indicator in the second step of our analysis (the dynamic network). Figure 5 shows these interactions for network pairs. Expectedly, the frequency of interactions for the regular-regular pairs outperformed the regular-ABIs pairs, which is reasonable due to the number of regular users. We can see that a similar linear pattern is obtained in both pairs. Moreover, we tested the variations of these interactions to see the stability of these interactions for each week. Figure 6 a) shows that the regular-regular pairs reflected a close level of variations compared to regular-ABIs pairs. In fact, this result needs more investigation in terms of its significance. Therefore, we decided to plot the statistical notched boxplot for the variation of both types of pairs. In Figure 6 b), the two boxes' notches do not show an overlap. This can be considered as strong evidence that their medians differ and this difference is statistically significant.

According to the obtained results, for the static network, it can be concluded that there exist interactions between regular and ABIs users, which lead the privacy of regular users to be exposed to the ABIs. Based on the results [28], similar behavior was obtained for the ABIs users.

The second step of our analysis is to evaluate the privacy issue in the dynamic network. To do this, we calculate the weight of all the pairs of regular-regular and regular-ABIs. We compare the level of interactions among users within the dynamic environment. More precisely, we try to find out how much the ABIs interacted with the regular users in terms of the features of encounters (recurrence, consistency, and duration). We believe that these indicators can reveal the strength of the relation of network pairs, which eventually lead to figuring out how much the privacy of regular users is exposed to the ABIs. Figure 7 depicts the weights of the two types of pairs. We observed that the weights of regular-ABIs pairs were growing faster than the regular-regular pairs. This is very interesting and promising since it leads to important findings. The relations of the regular-ABIs users became stronger as

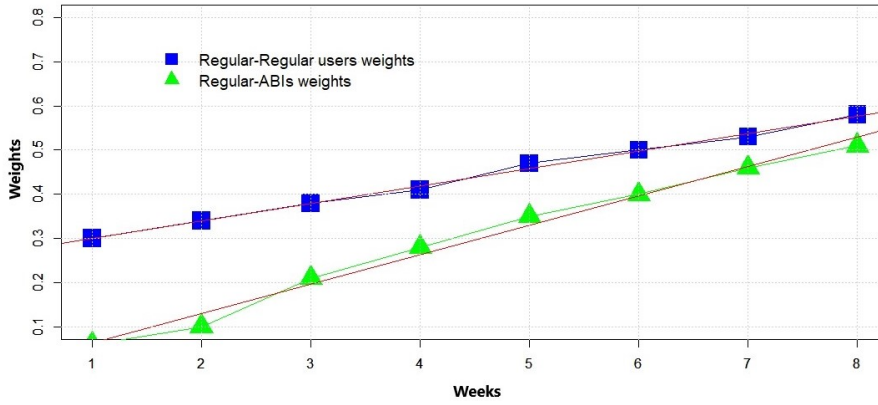


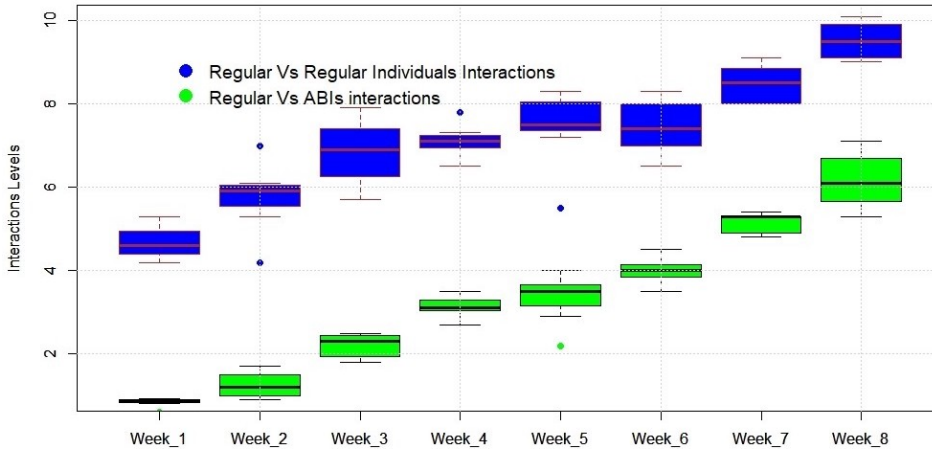
Figure 5. The interactions between regular-regular and regular-ABIs pairs in the static network. The x-axis represents the time from week 1 to 8 and y-axis the frequency of interactions.

time passes. This phenomenon could be considered as a negative indicator of privacy violations among users. In other words, the privacy of regular users is more exposed by the ABIs over time. According to these observations, it is needed to analyse the obtained results.

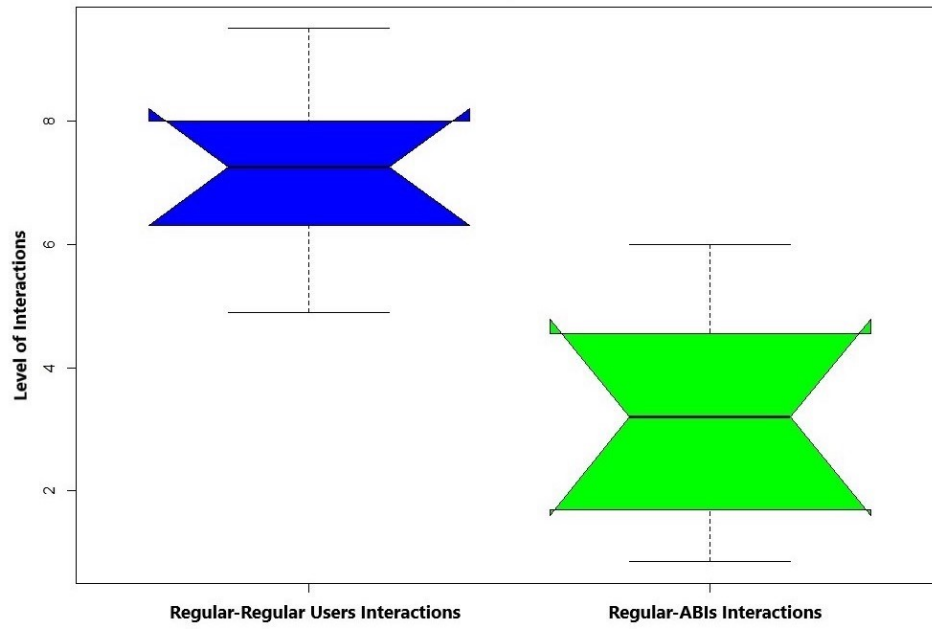
We start with the variations of the weight values of both types of pairs. Figure 8 shows the variations in weight for every week and for all the pairs. Figure 8 a) reflects the variations of every week and how they are changed and become close over time. Figure 8 b) shows the notches of both types of pairs, it is clear that their median is not overlapped, which means the difference has existed.

These results were not enough to be confirmed since after week_5 the pattern became closer. Therefore, it is important to investigate what driving this pattern and whether there is a statistically significant difference between both pairs. To confirm this, we create a linear regression model and then use one way ANOVA to analyse the variance of the pairs. As we can see from Table 1, the standard error of the interactions of regular-ABIs is less than regular-regular. Therefore, our *null hypothesis* ($H0$) assumes that the mean values of regular-regular and regular-ABIs are equal and the *alternative hypothesis* ($H1$) assumes not with a confidence level of 97 %. According to Table 1, the significance level is significantly greater than the *p-value*. This means we cannot accept the null hypothesis of equal means and the difference between both kinds of pairs is statistically significant. Also, Table 2 shows that the sum of squared errors for the regular-ABIs is significantly less than the other pair.

We can conclude that the interactions between regular and ABIs grow faster among regular users and the privacy of regular users is more exposed to the ABIs over time.



a)



b)

Figure 6. a) The variations of the interactions among regular-regular and regular-ABIs pairs in the static network. b) Notched Boxplot for the interactions of both classes of pairs.

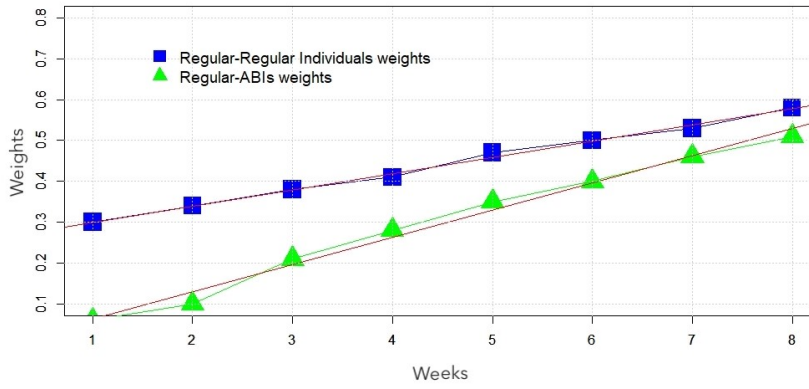


Figure 7. The interactions between regular-regular and regular-ABIs pairs in the dynamic network. The x-axis represents the time (week 1 to 8) and y-axis the weights.

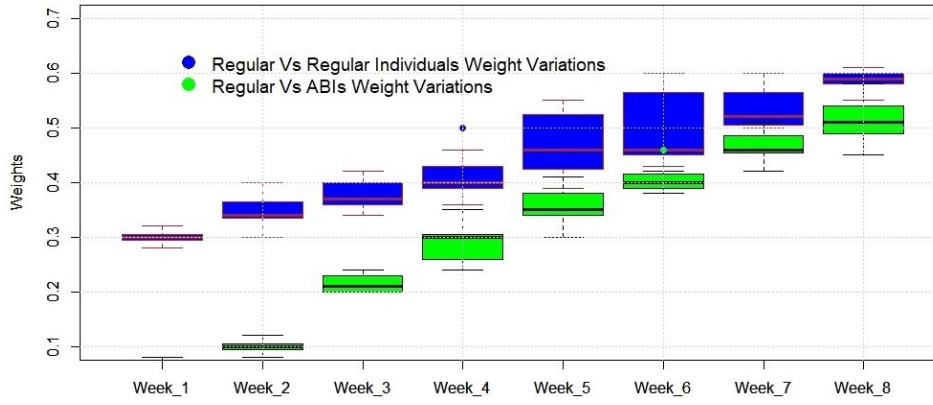
Coefficient	Estimate	Std. Error	P-Value	F-Statistics
Intercept	−4.886	1.42		
Regular-Regular Interactions	18.868	5.352	0.00000641	747.4
Regular-ABIs Interactions	3.737	3.176		

Table 1. One way ANOVA for regular-regular and regular-ABIs interactions

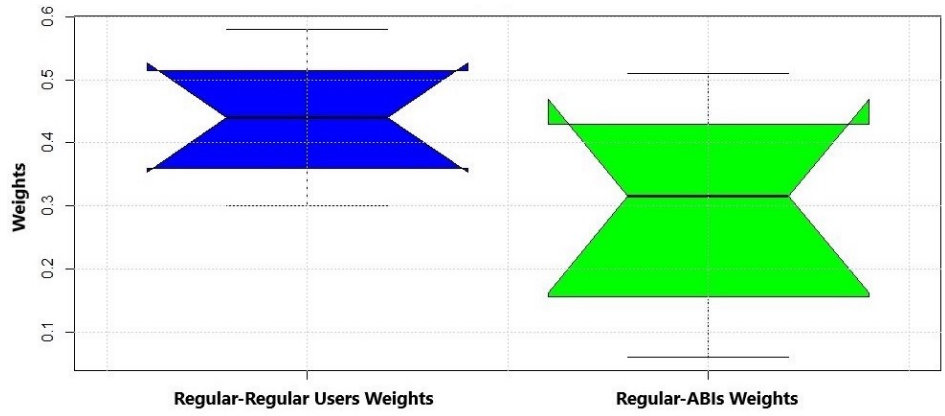
Another step in our analysis is investigating the correlations among the weekly behavior of the pairs in both networks (see Figure 9). This analysis is not related to the privacy issue insofar as it relates to the cumulative weakly behavior of users. The purpose of this analysis is to show how the behavior of users is changed from a week to another and whether there is a pattern that characterizes the correlations. In the static network, as shown in Figure 9 a), the weekly behavior of regular-regular users is close to a sine wave. The pairs ((week_2, week_3), (week_3, week_4), (week_5, week_6), and (week_6, week_7)) are strongly correlated. Since the dataset is related to demonstrations, we believe these correlations are due to the motivation of the demonstrators as well as the governmental procedures during the demonstrations. Moreover, the impact of ABIs on the privacy of users is cumulative, as can be seen in Figure 9 b). The figure shows that week 8 has correlations to approximately all the other weeks. This result confirms what we have seen in Figure 6 a) and the impact of ABIs is gradually increased and violated the privacy of regular users. On the other hand, Figures 9 c) and 9 d) show the correlations in the dynamic

Terms	Regular-Regular Inter.	Regular-ABIs Inter.	Residuals
Sum of Squares	41.821	0.038	0.639
Residual Std. Error		0.617	

Table 2. Sum of squares and residuals for regular-regular and regular-ABIs pairs



a)



b)

Figure 8. a) The variations of the interactions among regular-regular and regular-ABIs pairs in the dynamic network. b) Notched Boxplot for the interactions of both types of pairs.

network. The regular-ABIs pairs in Figure 9 d) do not reflect a clear weekly pattern because the movement patterns of users within the environment are different from regular users. More precisely, it is a consequence of the change in the exploration mechanism in the IM model. Also, we observe that regular-ABIs show different weekly pattern. For instance, the correlations between week 8 and the previous weeks are significantly weak compared to the correlations among the other weeks. This result interprets what has been mentioned in Figure 9 a) in week_8. Therefore, in dynamic networks, the correlation between two weeks is independent of the other pairs of weeks. Furthermore, the correlation between a week and its succeeding for the regular-ABIs pairs in both networks is shown in Figure 10. We can see

that the correlation between every two sequential weeks for the period of 8 weeks is unstable and significantly decreased over time. Since the features of the users in the dynamic network are inherited from the static, we observe that the pattern of the correlation is also reflected from the static network. It can be concluded that the privacy issue in online social networks has an approximately close pattern of exposure.

Based on the experience obtained from this work, our model can be improved – in terms of patterns' accuracy of the regular and ABIs users – by including more information and attributes about network users. However, the data collection process was restricted only to public data due to Facebook privacy policies that allows scrapers to collect only users' public information. On the other hand, incorporating more attributes and parameters into our dynamic model will increase the complexity of the model. This leads the dynamic simulations to generate a large-scale data that is not easy to handle. Furthermore, in case more information is available about users, some social concepts can be utilized such as the Homophily principle. This principle can be used in having a more in-depth investigations about users interactions (e.g., regular and ABIs).

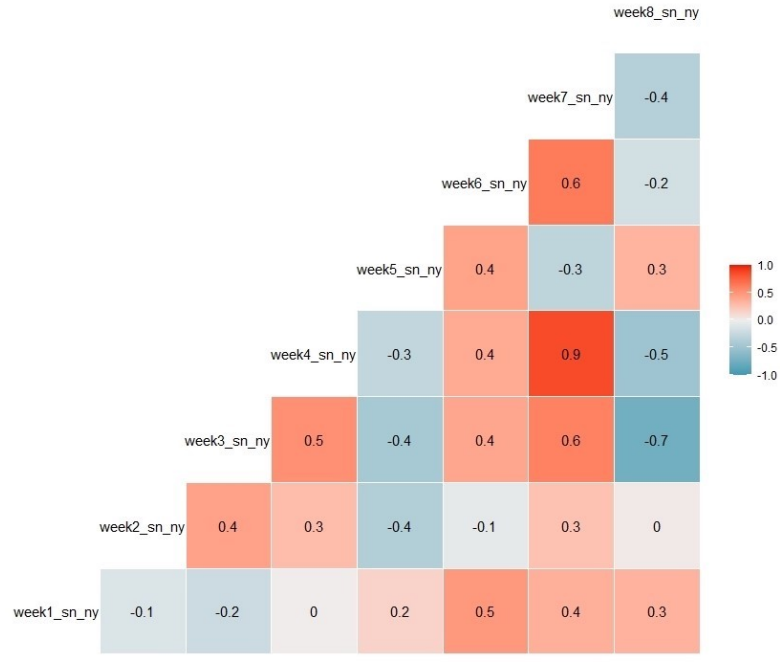
4 CONCLUSIONS

This article investigated future potential privacy violations in a dynamic network. The proposed approach projected a large scale static social network into a dynamic environment. The projection was not performed only based on users' relations, but also on their structural and spectral features. Therefore, the simulated dynamic network included the history of interactions of users in their online social networks.

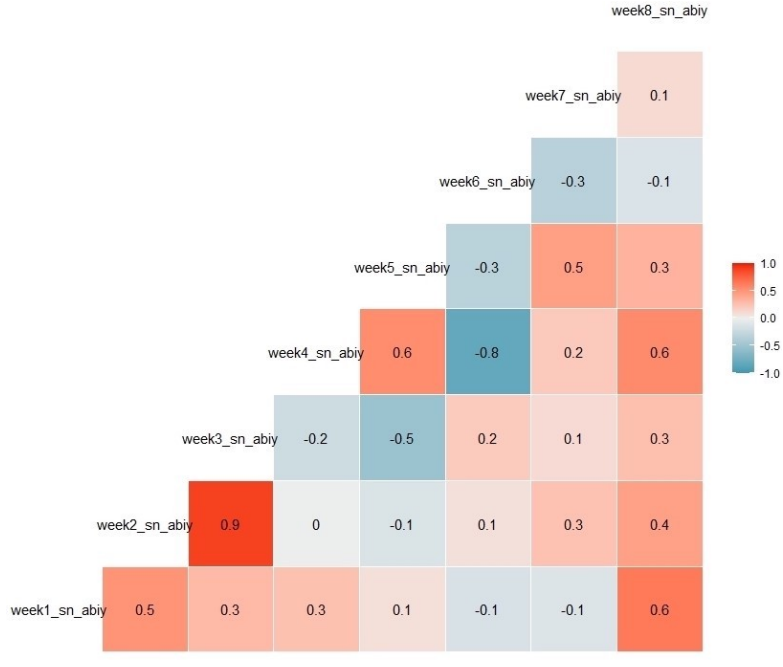
In the analysis, we involved two types of users: regular users and anomalous users. The results showed that violating the privacy of people in real life can be a side effect of their relations and interactions in online social networks. In real life, the level of violation can be increased over time due to the high possibility of interactions with anomalous users. Also, the interactions of regular people with anomalous in an online social network lead to exposing their privacy more than in dynamic networks. This is because people in real life are more aware of whom they are interacting with.

The main challenge in this research was the time consumed in running the simulations since we dealt with 27 835 dynamic nodes. The dynamics (e.g., interactions) of these nodes generated a significantly huge amount of data that should be keeping tracked for every single node within the simulation environment, which was not an easy task.

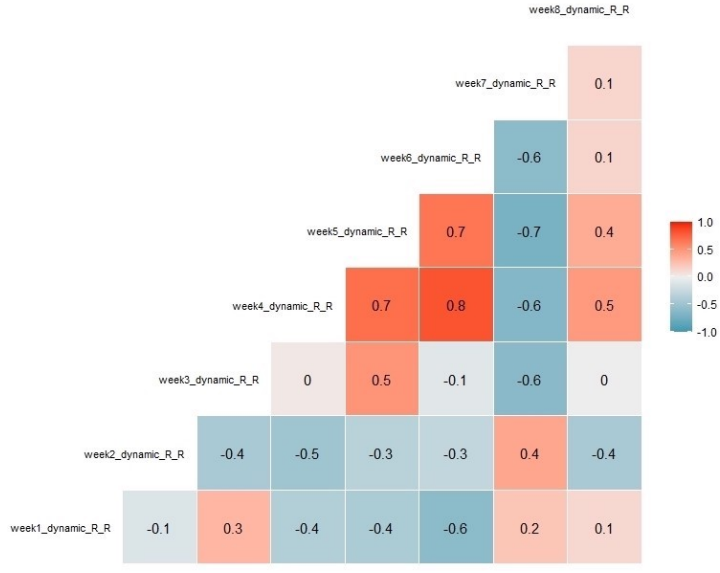
As future work, we plan to simulate our dataset for a longer time period (e.g., 3 months or more) and see the impact of users' interactions on their privacy. We also plan to reduce the dimensionality of features aiming at minimizing the simulations time. Other future works can be in adopting some principles in sociology


week8_sn_ny
week7_sn_ny -0.4
week6_sn_ny 0.6 -0.2
week5_sn_ny 0.4 -0.3 0.3
week4_sn_ny -0.3 0.4 0.9 -0.5
week3_sn_ny 0.5 -0.4 0.4 0.6 -0.7
week2_sn_ny 0.4 0.3 -0.4 -0.1 0.3 0
week1_sn_ny -0.1 -0.2 0 0.2 0.5 0.4 0.3
1.0
0.5
0.0
-0.5
-1.0

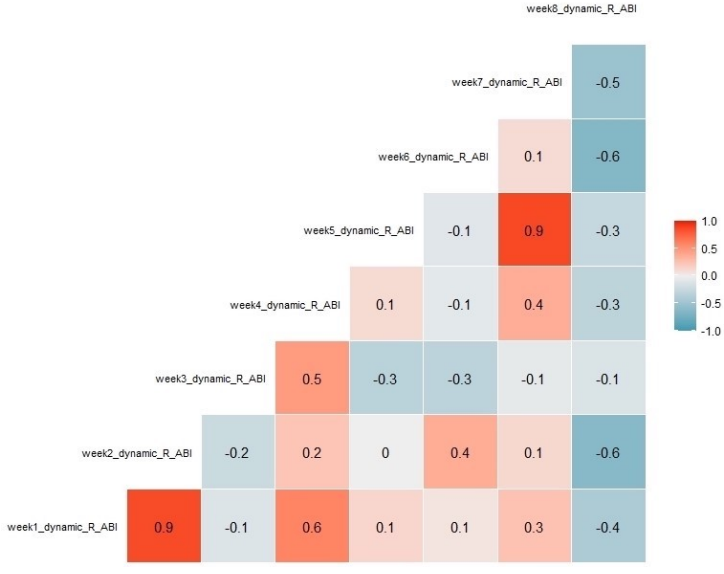
a)


week8_sn_abiy
week7_sn_abiy 0.1
week6_sn_abiy -0.3 -0.1
week5_sn_abiy -0.3 0.5 0.3
week4_sn_abiy 0.6 -0.8 0.2 0.6
week3_sn_abiy -0.2 -0.5 0.2 0.1 0.3
week2_sn_abiy 0.9 0 -0.1 0.1 0.3 0.4
week1_sn_abiy 0.5 0.3 0.3 0.1 -0.1 -0.1 0.6
1.0
0.5
0.0
-0.5
-1.0

b)



c)



d)

Figure 9. a) Static network correlations among regular-regular users. b) Static network correlations among regular-ABIs. c) Dynamic network correlations among regular-regular pairs. d) Dynamic network correlations among regular-ABIs pairs.

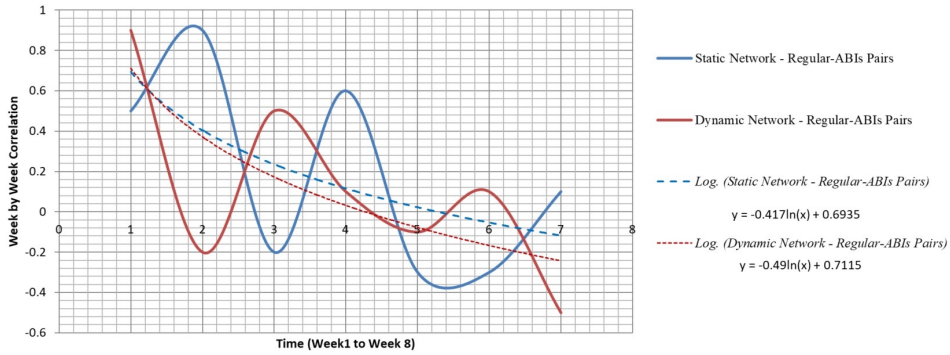


Figure 10. Correlations of the behavior for the pairs of weeks (week_1, week_2), (week_2, week_3), (week_3, week_4), (week_4, week_5), (week_5, week_6), (week_6, week_7), and (week_7, week_8) for both networks

such as Homophily (or called Assortativity) aiming at having an in-depth analysis of the regular-regular, ABIs-ABIs, and regular-ABIs interactions considering most of the possible attributes and parameters that make the simulations manageable.

4.1 Acknowledgement

The authors would like to thank the Computer Science Department, College of Computer Science and Mathematics, University of Mosul, Iraq, for all the support provided in making this work achieved.

REFERENCES

[1] Kim, J.—Hastak, M.: Social Network Analysis: Characteristics of Online Social Networks after a Disaster. International Journal of Information Management, Vol. 38, 2018, No. 1, pp. 86–96, doi: 10.1016/j.ijinfomgt.2017.08.003.

[2] Jaradat, S.—Dokoohaki, N.—Matskin, M.—Ferrari, E.: Trust and Privacy Correlations in Social Networks: A Deep Learning Framework. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2016, pp. 203–206, doi: 10.1109/ASONAM.2016.7752236.

[3] Kafali, Ö.—Günay, A.—Yolum, P.: Detecting and Predicting Privacy Violations in Online Social Networks. Distributed and Parallel Databases, Vol. 32, 2014, No. 1, pp. 161–190, doi: 10.1007/s10619-013-7124-8.

[4] Alrayes, F. S.—Abdelmoty, A. I.—El-Geresy, W. B.—Theodorakopoulos, G.: Modelling Perceived Risks to Personal Privacy from Location Disclosure on Online Social Networks. International Journal

of Geographical Information Science, Vol. 34, 2020, No. 1, pp. 150–176, doi: 10.1080/13658816.2019.1654109.

[5] Cho, J. H.—Alsmadi, I.—Xu, D.: Privacy and Social Capital in Online Social Networks. 2016 IEEE Global Communications Conference (GLOBECOM), IEEE, 2016, pp. 1–7, doi: 10.1109/GLOCOM.2016.7842168.

[6] Kökciyan, N.—Yolum, P.: PriGuard: A Semantic Approach to Detect Privacy Violations in Online Social Networks. IEEE Transactions on Knowledge and Data Engineering, Vol. 28, 2016, No. 10, pp. 2724–2737, doi: 10.1109/TKDE.2016.2583425.

[7] Savage, D.—Zhang, X.—Yu, X.—Chou, P.—Wang, Q.: Anomaly Detection in Online Social Networks. Social Networks, Vol. 39, 2014, No. 1, pp. 62–70, doi: 10.1016/j.socnet.2014.05.002.

[8] Greene, D.—Doyle, D.—Cunningham, P.: Tracking the Evolution of Communities in Dynamic Social Networks. 2010 International Conference on Advances in Social Networks Analysis and Mining, IEEE, 2010, pp. 176–183, doi: 10.1109/ASONAM.2010.17.

[9] Bindu, P.—Thilagam, P. S.: Mining Social Networks for Anomalies: Methods and Challenges. Journal of Network and Computer Applications, Vol. 68, 2016, pp. 213–229, doi: 10.1016/j.jnca.2016.02.021.

[10] Liao, Q.—Li, T.—Blakely, B. A.: Anomaly Analysis and Visualization for Dynamic Networks Through Spatiotemporal Graph Segmentations. Journal of Network and Computer Applications, Vol. 124, 2018, pp. 63–79, doi: 10.1016/j.jnca.2018.09.016.

[11] Chen, S.—Fu, A.—Shen, J.—Yu, S.—Wang, H.—Sun, H.: RNN-DP: A New Differential Privacy Scheme Base on Recurrent Neural Network for Dynamic Trajectory Privacy Protection. Journal of Network and Computer Applications, Vol. 168, 2020, Art. No. 102736, doi: 10.1016/j.jnca.2020.102736.

[12] Bhagat, S.—Cormode, G.—Krishnamurthy, B.—Srivastava, D.: Prediction Promotes Privacy in Dynamic Social Networks. Proceedings of the 3rd Wonference on Online Social Networks (WOSN '10), USENIX Association, 2010, pp. 1–9.

[13] Bhagat, S.—Cormode, G.—Krishnamurthy, B.—Srivastava, D.: Privacy in Dynamic Social Networks. Proceedings of the 19th International Conference on World Wide Web (WWW '10), ACM, 2010, pp. 1059–1060, doi: 10.1145/1772690.1772803.

[14] Farine, D. R.: When to Choose Dynamic vs. Static Social Network Analysis. Journal of Animal Ecology, Vol. 87, 2018, No. 1, pp. 128–138, doi: 10.1111/1365-2656.12764.

[15] Lopez, M.: Elite Theory. Sociopedia.isa, 2013, pp. 1–12.

[16] Higley, J.: Elite Theory and Elites. In: Leicht, K. T., Jenkins, J. C. (Eds.): Handbook of Politics: State and Society in Global Perspective. Springer, New York, 2010, pp. 161–176, doi: 10.1007/978-0-387-68930-2_9.

[17] Patterson, N.—Price, A. L.—Reich, D.: Population Structure and Eigenanalysis. PLoS Genetics, Vol. 2, 2006, No. 12, Art. No. e190, doi: 10.1371/journal.pgen.0020190.

[18] Newman, M. E. J.: Power Laws, Pareto Distributions and Zipf's Law. Contemporary Physics, Vol. 46, 2005, No. 5, pp. 323–351, doi: 10.1080/00107510500052444.

[19] Webb, W. R.—Gamsu, G.—Speckman, J. M.—Kaiser, J. A.—Federle, M. P.—Lipton, M. J.: Computed Tomographic Demonstration of Mediastinal Venous Anomalies. American Journal of Roentgenology, Vol. 139, 1982, No. 1, pp. 157–161, doi: 10.2214/ajr.139.1.157.

[20] Albert, R.—Barabasi, A. L.: Statistical Mechanics of Complex Networks. Reviews of Modern Physics, Vol. 74, 2002, No. 1, pp. 47–97, doi: 10.1103/RevModPhys.74.47.

[21] Smelser, N. J.: Theory of Collective Behavior. Quid Pro Books, 2011.

[22] Matsueda, R. L.: Cultural Deviance Theory: The Remarkable Persistence of a Flawed Term. Theoretical Criminology, Vol. 1, 1997, No. 4, pp. 429–452, doi: 10.1177/1362480697001004002.

[23] Stahel, W.—Limpert, E.: The Normal Distribution Is the Log-Normal Distribution. DIC-2014, 2014.

[24] Song, C.—Koren, T.—Wang, P.—Barabasi, A. L.: Modelling the Scaling Properties of Human Mobility. Nature Physics, Vol. 6, 2010, No. 10, pp. 818–823, doi: 10.1038/nphys1760.

[25] Grossman, A.—Sutton, J. R.: Endorphins: What Are They? How Are They Measured? What Is Their Role in Exercise? Medicine and Science in Sports and Exercise, Vol. 17, 1985, No. 1, pp. 74–81.

[26] Barabasi, A. L.: The Origin of Bursts and Heavy Tails in Human Dynamics. Nature, Vol. 435, 2005, No. 7039, pp. 207–211, doi: 10.1038/nature03459.

[27] Alanezi, M.—Mahmood, B.: Projecting Social Networks in Dynamic Environments for Tracking Purposes. 2021 2nd International Conference on ICT for Rural Development (IC-ICTRuDev), IEEE, 2021, pp. 1–5, doi: 10.1109/IC-ICTRuDev50538.2021.9655711.

[28] Mahmood, B.—Alanezi, M.: Structural-Spectral-Based Approach for Anomaly Detection in Social Networks. International Journal of Computing and Digital Systems, Vol. 10, 2021, No. 1, pp. 343–351, doi: 10.12785/ijcds/100134.

Mafaz Alanezi works as Associate Professor in the Computer Science Department at the University of Mosul, Iraq. Her Ph.D. and M.Sc. degrees were received from the University of Mosul. Her research areas of interest are cybersecurity and privacy, social media analysis, and mobile computing.

Basim Mahmood works as Associate Professor in the Computer Science Department at the University of Mosul. He is also the Member of the BioComplex Laboratory, Exeter, UK. His Ph.D. is from the Florida Institute of Technology, USA and his M.Sc. degree from the University of Mosul. His research interests are ubiquitous, pervasive and mobile computing, network science, and big data analysis.