Exploring patient-patient interactions graphs by network analysis

Zaher Salah[1], Esraa Abu Elsoud[2], Kamal Salah[3]

[1]Department of Information Technology, Faculty of Prince Al-Hussein bin Abdullah II for Information Technology,
The Hashemite University, Zarqa, Jordan
[2]Department of Computer Science, Faculty of Information Technology, Zarqa University, Zarqa, Jordan
[3]Deanship of Preparatory Year and Supporting Studies, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Article Info	ABSTRACT
	Understanding how patient demographics and shared experiences impact interactions is essential for strengthening pa/tient support networks and optimizing health outcomes as personalized healthcare becomes more and more important. To this end, this study explores the patient-patient interactions (PPIs) graph as a network and applies selected network analysis approaches to examine the PPIs network of accutane drug. Two main research questions are addressed by gaining deeper insight at the hidden patterns of reactivity and connectivity among interchanging nodes. There was a negative response to the first research question, which asked if patients react to others that have similar gender and/or age profiles in a consistent way. Patients tended to interact with people of different genders and ages, indicating a high degree of heterogeneity in the network. Negative responses were likewise given to the second research question, which asked if communities inside the network could identify patients based on gender or age profile. Network analysis approaches for community detection failed to distinguish between groups with similar demographic characteristics. Rather, groups seemed to emerge based on other factors, like similarity in patient opinions. The results imply that gender and age do not have a major influence on community membership. Future research will concentrate on applying more sophisticated graph mining techniques to expand these approaches to cover more and larger PPIs networks.

Corresponding Author:

Zaher Salah
Department of Information Technology
Faculty of Prince Al-Hussein bin Abdullah II for Information Technology, The Hashemite University
P.O. Box 330127, Zarqa 13133, Jordan
Email: zaher@hu.edu.jo

1. INTRODUCTION

A new heterogeneous network embedding technique called self-data heterogeneous information network embedding (SDHINE), which incorporates patient-patient interactions (PPIs) data into drug embeddings and is applicable to various kinds of adverse drug reaction (ADR) prediction tasks, was described by Baofang *et al.* [1]. The authors first designed various meta-path-based proximities to calculate drug similarities, particularly target propagation meta-path-based proximity based on PPI network, and then built a semi-supervised stacking deep neural network model that is jointly improved by the defined meta-path proximities in order to integrate mixed drug information and learn drug representations. The efficacy of the SDHINE model is proven by comprehensive evaluations on three ADR prediction tasks using three modern

network embedding techniques. Additionally, by mapping the drug representations into 2D simpler space, the authors compared the drug representations in terms of drug discrimination. The results demonstrated that the proposed technique performed better than the comparative methods. Zhao *et al.* [2] used the network embedding technique known as Mashup in their research to extract important and informative drug features from a number of drug heterogeneous networks that represented various pharmacological features. In order to extract side effect features, a network was also constructed for side effects. These functions are capable of gathering crucial data at the network level on drugs and their adverse side effects. Each pair of drug and side effect was represented by combining aspects of the drug and the adverse effect. Moreover, they were input into the random forest (RF) network model, a prediction model created by the RF algorithm. Following several rounds of tests, the average Matthews correlation coefficients for the balanced and unbalanced datasets were found to be 0.640 and 0.641, respectively, according to the experimental results evaluating the RF network model. Compared to earlier models using other machine learning algorithms, the RF network model performed better.

A new approach to predicting possible drug side effects was established in the research work described in [3]. This approach is based on more complete information about drugs that integrates the drug's forms of effect on proteins of interest. A certified heterogeneous information network is used to model several sorts of drug information. Using two bias random walk methods to extract drug sequences and train a skip-gram model to learn drug embedding, the authors presented a verified heterogeneous information network embedding framework for learning drug embedding and predicting drug side effects. By contrasting the outcomes of the experiments with the most advanced techniques, the proposed method's performance was proved. Moreover, a case study's outcomes validate the hypothesis that drugs effects on targeted proteins are beneficial for side effect predicting. Yang and Zhao [4] developed a systematic method that uses online health communities (MedHelp) and pharmaceutical repositories (PharmGKB and SIDER) to identify repositioning drugs through heterogeneous network analysis. The authors created a heterogeneous health network comprising drugs, diseases and ADRs by using ADRs as the intermediary. They also created path-based heterogeneous network mining techniques for drug repositioning.

Additionally, they looked into how the effectiveness of drug repositioning is impacted by the information sources. The outcomes of the experiment shown that merging PharmKGB and MedHelp offered 479 repositioning drugs more than the number of repositioning drugs discovered through other approaches. Furthermore, PubMed data aided 31% of the 479 repositioning drugs that were discovered. A new computational methodology known as graph attention-based convolutional learning for CircRNA-disease prediction (GATCL2CD) was presented in [5] in order to predict unidentified circRNA-disease associations (CDAs). Gaussian interactive profile kernel (GIP) similarity and semantic similarity for illnesses, circRNA sequence similarity and function similarity, and GIPs for circRNAs were first computed by the authors. They then joined them together to create a heterogeneous graph. After that, the feature convolution machine learning model GATCL2CD was developed. It generated various aggregated representations of features that related to the nodes in the heterogeneous graph with the assistance of a multi-head dynamic attention approach. A single-layer convolutional neural network employing filter kernels of various sizes was then used to extract better higher-order attributes from each node's stacked attribute representations. In the end, a multi-layer perceptron neural network was shown as an effective classifier to predict possible CDAs, and a pairwise element-wise product operation was established to identify the interactions of higher-order attribute representations. Solid experimental findings on three distinct datasets using 5-fold cross-validation shown that GATCL2CD outperformed five new approaches. Additionally, case studies proved that GATCL2CD is a good tool for discovering possible circRNAs linked to diseases. PrimeKG, a multimodal knowledge graph for precision drug analysis, was proposed by the authors in [6]. PrimeKG significantly expanded previous efforts in disease- associated knowledge graphs by integrating 20 outstanding resources that characterize 17,080 diseases with 4,050,249 relationships representing ten major biological scales: disease-associated protein perturbations, biological processes and pathways, anatomical and phenotypic scales, and an extensive list of approved drugs with their therapeutic effect. PrimeKG can facilitate AI investigations of how pharmaceuticals affect disease-associated networks since it has an extensive number of "indications," "contradictions," and "off-label use" drug-disease edges that are unavailable in other knowledge graphs.

2. METHOD
2.1. Patient-patient interactions graph (network)

In the research conducted for this paper, patients and caregivers submitted textual patient reviews, which were published online in HTML format (at www.druglib.com), with a primary focus on the drug side effects section. Because it offers comprehensive, organized and up-to-date drug information including side effects, effectiveness and individual responses from patients, data from www.druglib.com was utilized. Because it frequently originates from clinical investigations and authorized organizations the data is reliable

where this is crucial for assuring the accuracy and practicality of analyses in research in medicine. With respect to patient's privacy, approaches like the integrated health profile (IHP) proposed by [7] can be adopted. IHP is a decentralized and impermeable platform for safely storing and exchanging medical records that makes use of smart contracts and blockchain technologies. Every patient's IHP card, which contains all medical records like reports, prescriptions and bills, is assigned with a unique identifier. Medical practitioner can scan the QR code on the IHP card to start a two-phase authentication procedure that asks the patient to enter a one-time password (OTP). By limiting access to certain shared records, this authentication protects patient privacy and guarantees safe and monitored data access. Salah *et al.* [8] used their framework to extract the necessary data about nodes, links, and required labels in order to create the associated PPIs graph. The high-level structure of these dissatisfactions could be graphically visualized by using sentiment analysis techniques to extract (PPIs) graphs. This allowed for a more complete comprehension of the information hidden in a large quantity of these written representations for the corresponding patients' reviews. The principle behind this approach is to use a graph to visualize the interactions' structure, with patients participating as nodes and interactions as links. Based on the text similarity of nodes, links are created. SentiWordNet 3.0 sentiment lexicon is used to classify nodes based on the patient's attitude toward a certain drug, whether it be positive or negative. Next, attitudes are used to classify the graph linkages as either in favor of or against drug use. If the two patients have the same attitude that is, a negative attitude regarding severe side effects or a positive attitude regarding moderate side effects the relationship is deemed supportive; if not, it is deemed opposing. The consequent graphs show drugs as the subject of a disagreement between two opposing groups. The PPIs graph extraction methodology is illustrated in Figure 1.
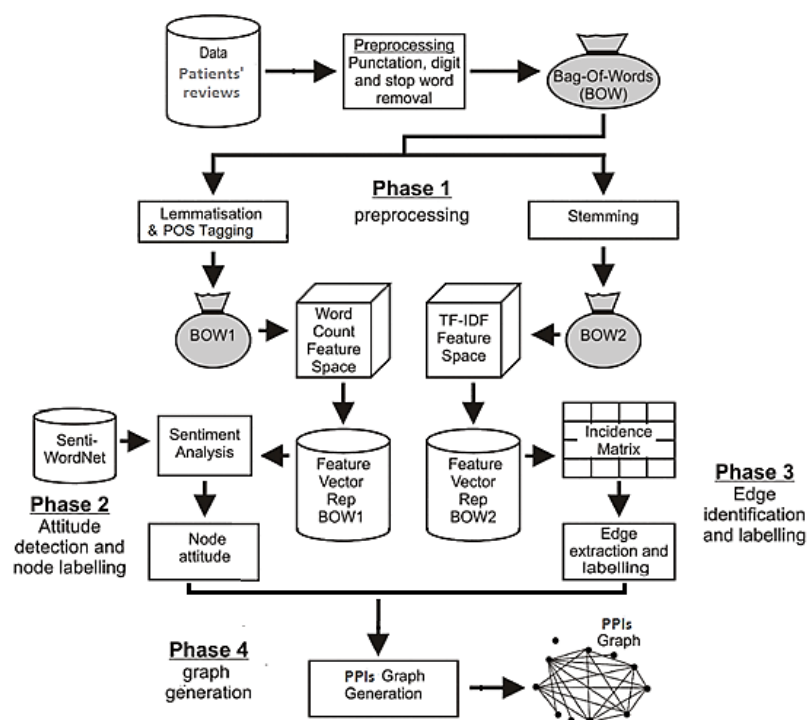

Figure 1. PPIs graph extraction framework

Accutane (isotretinoin), one of the drugs from our DrugLib patient reviews dataset, is used to generate PPIs graphs. The resulting graph is presented in Figure 2. Figure 2 shows a designated node for each patient, labeled with the patient's age (the number between parentheses) and gender (F: female, M: male). From the patient's perspective on the drug under consideration, a green node indicates a patient with a positive attitude (moderate side effect) and a red node indicates a patient with a negative attitude (severe side effect). When two linked nodes represent two patients, the thickness of the link between them reflects how similar their semantic material is. This is calculated by summing up all the phrases (words) in both reviews that have non-zero weights for term frequency-inverse document frequency (TF-IDF) and appear to be about the same issue. The links with green colors indicate people who endorse or are in approval, while the links with red colors indicate those who are against.
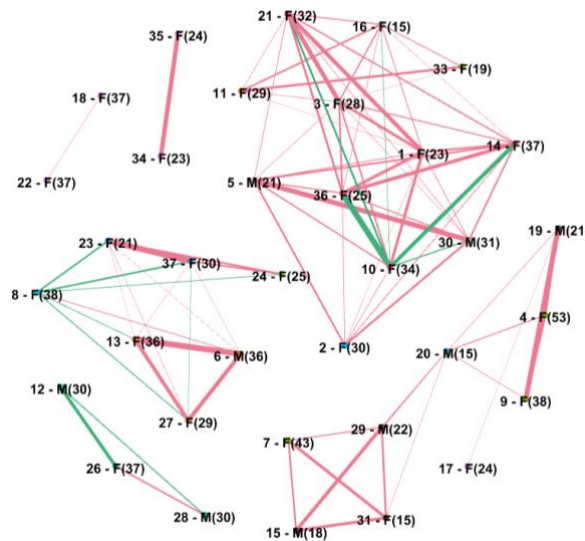
The visualization was produced using Gephi at https://gephi.org/

Figure 2. PPIs graph for accutane drug (isotretinoin) produced from druglib.com patients reviews

The authors will use the information mentioned earlier to demonstrate how PPIs graphs can be used to facilitate various levels of analysis. PPIs graphs can be used, in more detail, to examine: i) if patients consistently responded to those who did fit the same gender or age profile and ii) if the PPIs graph (network) "community" reflect a gender or age profile. The core objective of the research work presented in this paper is to identify the structural properties and highlight some of the features of the graphs, such as how patients are likely to rank the side effects of drugs, how patients interact in their reviews, and which patients are more influential, by applying network analysis techniques to the graphs see the examples on combining sentiment analysis and networks analysis presented in studies [9]–[17]. To the best of the author's knowledge, no earlier research has made an attempt to characterize and analyze patient reviews in this way with a concentration on side effects in the context of a certain drug. The research described here aims to analyze the latent graph structures that are existent in graphs of PPIs in relation to the interactions between the individual patients. Is it possible to use applicable techniques from the field of network analysis to represent and analyze PPIs as graphs (conceptualized as networks)? More precisely, what network analysis metrics and methods should be applied to draw attention to the structural characteristics of these kinds of graphs?

The authors will explain an approach for analyzing PPIs graphs using network metrics and community detection algorithms in the section that follows. This approach is based on a pilot study. The importance of this research is based on the fact that by examining existing patterns of connections and involvement between the exchanging nodes (patients), network measurements and community detection computational methods can be used to anticipate outcomes.

2.2. PPIs graph analysis

This section explains how to effectively make use of (PPIs) graphs for supporting different types of analysis, as they are constructed using the (PPIs) graph extraction framework. PPIs graphs can be used in particular to: i) investigate whether patients consistently responded to other patients who had a similar gender or age profile; and ii) investigate whether the gender or age profile of the "community" inside the PPIs graph (network) is indicated. The first, it focuses on (PPIs) graphs and discusses the nature of the arguments between the two parties about an associated drug. The second deals with recognition of communities within (PPIs) graphs and the possible interpretations of these communities' characteristics. Since the theory of network analysis is the foundation of both types of investigations [18], (PPIs) graphs can be interpreted as networks. Clustering coefficient concept is recommended for the first type of analysis (assortativity can also be used [18]). Different network community detection computational methods can be successfully applied for the second type of analysis, as will be covered in more detail later in this section. With respect to the intended (PPIs) graph, the following exemplar questions were taken into consideration in order to demonstrate the usefulness of the graph in the context of the two types of analysis mentioned in section 2.3:

Q1: Are patients consistently responding to patients belonging to a similar gender and/or age profile?
Q2: Are communities found in the (PPIs) network able to identify, at least roughly, a patient's age or gender?

2.3. The accutane (isotretinoin) network

With 35 nodes representing each patient who participated in the drug reviewing and 86 edges representing patient interactions, the accutane drug's (PPIs) network (undirected graph) was constructed. Figure 3 shows the network's degree distribution, while Table 1 provides information on the accutane (isotretinoin) nodes. Highly connected nodes are fewer in number than poorly connected nodes, as is to be expected. From a network analysis perspective, it makes sense to investigate if a network's degree distribution fits a power-law distribution. The accutane network's degree distribution is right-skewed and roughly follows a power-law distribution, as shown by the histogram of degree distributions in Figure 3. Scale-free networks are defined as networks having degree distributions that follow a power law [18].

A subset of nodes from a graph connected by a path is referred to as a weakly connected component in graph theory terminology. Therefore, the process must locate every weekly connected component of the network in order to obtain the list of nodes that are part of the same cluster or group of overlapping clusters. This process is carried out as a depth-first search, which investigates a graph in its entire form, digging as far as possible into each of its branches before backtracking. With V representing the number of vertices or nodes and E representing the number of edges in the graph, its time complexity is $O\ (|V|+|E|)$. The nodes and edges for each weakly connected component are acquired by visiting every vertex in the graph [19]. Reconstructed clusters consist only of connected components that have more than one node. A connected component in static graphs is the largest possible set of vertices connected by graph edges. In simpler terms, if there is a path in the graph connecting two vertices, u and v, in the component, then it exists. Strongly and weakly linked components can be used to expand the concept of directed graphs in two different ways: either there is a directed path from u to v and one from v to u, or only one of those paths exists [20].
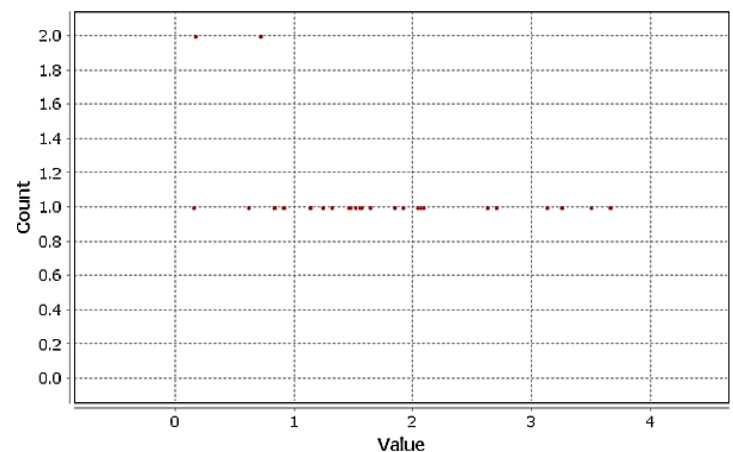

Figure 3. The accutane (isotretinoin) graph average weighted degree: 1.691

Table 1. The accutane (isotretinoin) nodes information

Graph elements	Statistical summary
Nodes	35
Edges	86
Average degree	4.914
Average weighted degree	1.691
Network diameter	4
Graph density	0.145
modularity	0.638
Average clustering coefficient	0.867
Average path length	1.398
■ Moderate side-effects	84.88%
■ Severe side-effects	15.12%
Connected components	6
Compnent-1	34.29%
Compnent-2	25.71%
Compnent-3	20%
Compnent-4	8.57%
Compnent-5	5.71%
Compnent-6	5.71%

2.4. Analysis of accutane network

In this research, PPIs graphs are examined using two forms of network analysis. To answer research question Q1, the clustering coefficient is first used. Second, betweenness centrality is used for community structures detection for answering research question Q2.

2.4.1. Clustering coefficient

A measure of how much nodes in a graph tend to cluster together is called a clustering coefficient in the context of graph theory. It measures the degree of cohesion in a node's neighborhood within a network. It is classified into two categories: local values, which quantify the cohesion surrounding a particular node, and global values, which quantify the clusters within the network as a whole. It should be underlined that only single-edge graphs can use both of the clustering coefficient's formulations. Additionally, many edges are not taken into consideration in the majority of measurements in real-world networks. They only take into account basic graphs free of loops and multiple edges as a result. For weighted graphs, the clustering coefficient can also be derived [21]–[23]. The ratio of edges neighboring nodes to all potential edges between them is known as the clustering coefficient. An average measurement of node clustering in a network is given by the global clustering coefficient. Stronger node tendency to form densely connected clusters is indicated by higher clustering coefficients. Prior studies have demonstrated that networks with random and scale-free characteristics typically have poor clustering coefficients. On the other hand, networks with larger clustering coefficients have proven to exhibit a higher level of correlation [24].

The probability that any two randomly selected neighbors of a vertex v, of degree at least 2, are linked together is known as the clustering coefficient of v. If $d(v)$ represents the number of neighbors of v, then the calculation is $\binom{d(v)}{2}$=number of triangles containing v divided by number of potential edges between its neighbors. The average of this value for all vertices of degree at least 2 in the graph may then be used to define the clustering coefficient of the entire graph [25]. Figure 4 shows the accutane (isotretinoin) graph clustering coefficient metric report (clustering coefficient distribution): parameters: network interpretation: undirected, results:average clustering coefficient: 0.867, total triangles: 130, the average clustering coefficient is the mean value of individual coefficients.
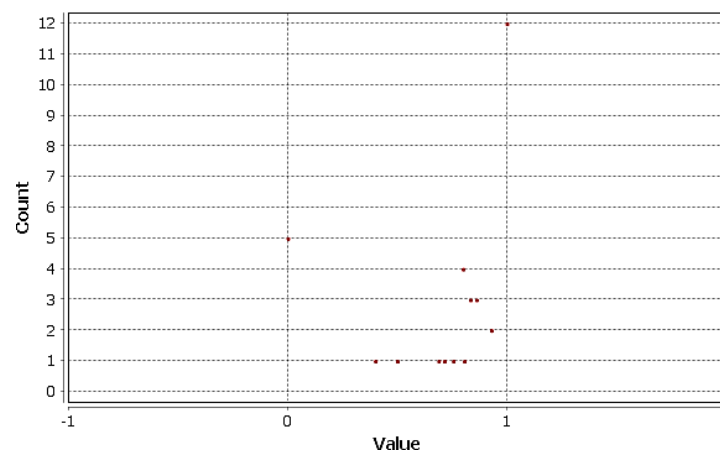

Figure 4. The accutane (isotretinoin) graph clustering coefficient metric report

2.4.2. Community structures detection

In social network analysis, the betweenness centrality index is crucial, although it is expensive to calculate. The least time-consuming methods available now take $O(n^2)$ space and $O(n^3)$ time, where n is the number of nodes in the network. The increasing demand for centrality measures on sparse, large-scale networks has led to the introduction of new betweenness algorithms in [26]. For unweighted and weighted networks, respectively, they take up $O(n+m)$ space and execute in $O(nm)$ and $O(nm+n^2 \log n)$ time complexity, where m is the number of links. This significantly broadens the variety of networks for which centrality analysis is practical, as demonstrated by experimental data. Centrality indices formed on graph vertices are a crucial tool for social network analysis. They are intended to represent the importance of nodes tangled in a social structure and are used to rank the nodes based on where they are in the network. Various

centrality indices, such as those that measure a node's average distance from other nodes or the ratio of shortest paths that a node lies on, are based on the shortest paths that link pairs of nodes.

An assessment of these indices is a fundamental component of many network-analytic research [26]. A network node's importance is measured by betweenness centrality [24], [27], which is based on shortest paths and reflects nodes' contributions to structural stability, social influence, and information diffusion. betweenness centrality is frequently used across various domains, including influence evaluation, community discovery, and social network analysis. The Brande's algorithm [26] is the most effective algorithm for calculating betweenness centrality quickly. It is based on the observation that the betweenness centrality value of a node v is equal to the total of all the fractions of shortest paths from other node pairs *(st)* that pass through node v. Using this formula as a starting point, the Brande's algorithm discovers the shortest paths between each node v and every other node, documenting the frequency and number of each node along the shortest paths. The betweenness centrality values of each node, starting with the leaf nodes and ending at the root node, are then summed up based on the information gathered. When calculating the betweenness centrality of every node in an unweighted graph, the Brande's algorithm needs *O(nm)* time complexity, where n is the number of nodes in the network and m is the number of edges. The Brande's algorithm operates with an $O(nm+n^2logn)$ time complexity for weighted graphs. Large-scale networks still find these time complexity to be prohibitive, consequently a reliable and effective betweenness centrality approximation algorithm is necessary. Let $G=(V,E)$ be a graph. G can be either directed or undirected, and the edge weights must be non-negative. $n=|V|$, $m=|E|$ and the number of shortest paths from node s to node t is represented by σ_{st}, while the number of shortest paths that pass via node v is represented by $\sigma_{st}(v)$. The betweenness centrality or BC value of a node $v \in V$ in a graph $G=(V,E)$ as shown in (1):

$$BC(v) = \sum_{\substack{s \neq t \neq v \\ s,t \in V}} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

Based on the Brande's algorithm's pair dependency, we can derive as shown in (2):

$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}} \ and \ \delta_{s^*}(v) = \sum_{\substack{t \neq v \\ t \in V}} \delta_{st}(v) = \sum_{v \in r_s(w)} \frac{\sigma_{sv}}{\sigma_{sw}} \cdot \left(1 + \delta_{s^*}(w)\right) \tag{2}$$

Where the set of all antecedents of node w is denoted by $r_s(w)$, we may recalculate the new betweenness centrality formula using (1) and (2) as shown in (3):

$$BC(v) = \sum_{\substack{s \neq v \\ s \in V}} \delta_{s^*}(v) \tag{3}$$

3. RESULTS AND DISCUSSION

The detailed summary of the numerical results of network analysis processes conducted on the accutane drug's PPIs network are presented in Table 2. This PPIs network contained 35 nodes representing the patients who participated in the drug reviewing and 86 edges representing patient interactions or semantic relationships between reviews made by those patients. Node attributes include node-ID, gender and age. We may gain deeper insight about the network's behavior and structure by examining the recorded findings of several important metrics that were provided. Each node (patient) in the PPIs network was represented by a row in the table. Label provided additional information about each node, including its unique identifier (node-ID), age and gender (for example, F(23) indicates a female node aged 23, and M(21) indicates a male node aged 21), degree (number of connections), weighted degree (strength of connections), betweenness centrality (a measure of node influence over information flow), and clustering coefficient (a measure of neighborhood network density). In addition, the component number (identifying the connected subnetwork to which the node belongs) and the number of triangles (groups of three connected nodes) were provided by the table.

The number of direct connections, or edges, that a node has with other nodes is counted by the degree metric. Higher degree nodes interact with other nodes in the network more frequently. Figure 5 shows the accutane (isotretinoin) graph distance report (betweenness centrality distribution). With a total degree of 172 and an average degree of 4.914 for all nodes, each node has roughly 5 connections on average. With an average of 1.691 and a total weighted degree of 59.192, it appears that the strength of the links varies. The two nodes with the highest degree in the table, nodes 1 and 3 F(23) and F(28), have 10 direct connections in the network, indicating that they are at the center of the network and interact with 10 other nodes. In contrast, Nodes 17 (F(24)) and a few others have only 1 degree, indicating minimal interaction. Weighted degree

enhances the degree measure by incorporating each connection's weight or strength instead of basically its count. Weights can be used to reflect a connection's importance, frequency, or intensity. For instance, node 10 (F(34)) has the largest weighted degree of 3.65 indicating that there are a lot of connections, but they are also stronger than those of other nodes. Although having a raw degree of 10, node 1 (F(23)) has a weighted degree of 3.49, indicating that its connections are not all that strong.

Table 2. The accutane (isotretinoin) summary of the numerical results

| Node-ID	Label	Degree	Weighted degree	Betweeness centrality	Component number	Clustering coefficient	Triangles
1	F(23)	10	3.49	3.61	1	0.76	34
2	F(30)	6	1.51	0.00	1	1.00	15
3	F(28)	10	3.12	6.46	1	0.69	31
4	F(53)	3	1.13	0.00	2	1.00	3
5	M(21)	8	2.69	0.42	1	0.93	26
6	M(36)	5	2.08	0.00	3	1.00	10
7	F(43)	3	1.13	0.00	2	1.00	3
8	F(38)	6	1.56	1.00	3	0.80	12
9	F(38)	3	1.45	0.00	2	1.00	3
10	F(34)	9	3.65	1.10	1	0.86	31
11	F(29)	5	1.31	0.76	1	0.80	8
12	M(30)	2	0.83	0.00	4	1.00	1
13	F(36)	5	2.03	0.00	3	1.00	10
14	F(37)	9	3.25	1.10	1	0.86	31
15	M(18)	4	1.63	1.67	2	0.83	5
16	F(15)	8	1.91	3.28	1	0.71	20
17	F(24)	1	0.15	0.00	2	0.00	0
18	F(37)	1	0.16	0.00	5	0.00	0
19	M(21)	4	1.54	7.00	2	0.50	3
20	M(15)	6	1.23	16.00	2	0.40	6
21	F(32)	9	3.24	1.75	1	0.81	29
22	F(37)	1	0.16	0.00	5	0.00	0
23	F(21)	6	2.05	1.00	3	0.80	12
24	F(25)	3	0.91	0.00	3	1.00	3
26	F(37)	2	0.90	0.00	4	1.00	1
27	F(29)	5	1.84	0.00	3	1.00	10
28	M(30)	2	0.61	0.00	4	1.00	1
29	M(22)	4	1.47	1.67	2	0.83	5
30	M(31)	8	2.62	0.42	1	0.93	26
31	F(15)	4	1.56	1.67	2	0.83	5
33	F(19)	3	0.82	0.00	1	1.00	3
34	F(23)	1	0.71	0.00	6	0.00	0
35	F(24)	1	0.71	0.00	6	0.00	0
36	F(25)	9	3.65	1.10	1	0.86	31
37	F(30)	6	2.06	1.00	3	0.80	12
	Total	172	59.192	51.000		26.005	390
	Average	4.914	1.691	1.457		0.743	11.143
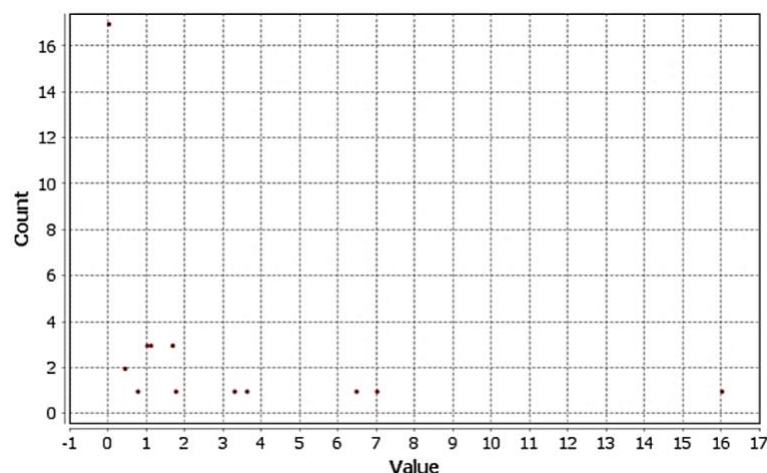
Figure 5. The accutane graph distance report (betweenness centrality distribution): parameters: network interpretation: undirected, results: diameter: 4, radius: 1, average path length: 1.3984375

The network has a total degree of 172 and a total weighted degree of 59.192. This implies the total connection strength for all of the nodes. With an average degree of 4.914, every node in the network interacts with roughly five other nodes on average. The clusters or sub-networks to which the nodes belong, or parts of the larger network, are referred to by their component numbers. Connectivity exists between nodes within the same component, but disconnectivity occurs between nodes located in separate components. It is evident that the network is not entirely connected because it consists of multiple disconnected components (numbered from 1 through 6). Although there are isolated nodes (like node 17 with component number 2), there are also large components like 2 and 3. For instance, node 1 and additional nodes like node 2 are part of component 1, indicating that together they constitute a coherent subnetwork. Node 12 is a member of component 4, which means it is a separate isolated cluster. Triangles, which are composed of three connected nodes, represent the number of triangle relationships that a node is a member of. Strong community structure is indicated by high triangle numbers. Node 1 is involved in 34 triangles, indicating a high number of three-way interactions. Nodes 17 and a few other nodes, on the other hand, do not form any triangles, highlighting their isolation or lack of community interaction. There are 390 triangles in total, and each node has an average of 11.143 triangles, indicating that nodes typically belong to small connected groupings that represent community-like interaction.

Betweenness centrality quantifies a node's importance for establishing interactions by calculating how far it is along the shortest paths connecting other nodes. A high betweenness centrality value indicates that the node serves as a network bridge within the network. The average betweenness centrality is 1.457, meaning that nodes have a moderate impact on establishing connections between other nodes. The total betweenness centrality is 51. As an illustration, node 3 (F(28)) has a considerable betweenness centrality of 6.46, reflecting its significance in connecting together the network's elsewhere separated sections. With a betweenness centrality of 16.00, node 20 (M(15)) has the highest betweenness and is therefore very important to the information flow across the network's communication structure. On the other hand, a large number of nodes have a betweenness centrality of 0, indicating that they are not central or act as outsiders. The clustering coefficient indicates the degree of local cohesiveness or cliquishness (producing a complete clique) by calculating the degree to which a node's neighbors are connected to one another. A cohesive community is indicated by a high clustering coefficient value. A clustering coefficient of 0.00 (as seen in node 17) indicates that the node's neighbors do not form any triangles, while a clustering coefficient of 1.00 (as in node 2) implies that all local neighbors are fully connected and contributing to a tightly bound (highly connected) cluster. The network's average clustering coefficient of 0.743 indicates that nodes are fairly clustered with many tightly connected groups and a high tendency for local clustering.

4.   CONCLUSION

This paper presents a study on the conceptualization of PPIs graph as a network and analyzing this network by means of selected network analysis approaches through the exploration of the hidden patterns of reactivity and connectivity among interchanging nodes. The PPIs network of Accutane drug was selected. The process of the analysis of this PPIs network was explained in detail and thus the objective for this study was addressed. It can be observed from the foregoing that: i) research question (Q1: Are patients consistently responding to patients belonging to a similar gender and/or age profile?) was answered negatively because the network exhibited high degrees of heterogeneity with respect to gender and/or age profile with different values. This emphasized a disagreement between gender and age profile. Patients tend therefore to interact with patients with different gender and/or age profile. ii) research question (Q2: Are communities found in the (PPIs) network able to identify, at least roughly, a patient's age or gender?) was answered negatively because none of the considered community detection approaches was able to detect communities, within the network, of members having the same gender or the same age profile. Some communities (e.g., component 1) contained nodes (patients) from different age and gender profiles. Other components (like 4 and 5) are much smaller and could represent more homogenous groups considering age or gender. Thus it was concluded that, the community structure (identified by component numbers) contain nodes with high clustering coefficients seemed not to be highly correlated with age or gender. These communities may be formed based on other factors like similarity between patients' reviews lexical contents or sentiments, rather than gender or demographic characteristics, i.e., age or gender was not a dominant factor in determining membership in these communities. Many promising future research directions present themselves, so as to extend the functionality and enhance the operation of analyzing large collections of PPIs networks directly using graph mining approaches rather than using basic tabular data analysis techniques.

FUNDING INFORMATION

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Zaher Salah	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Esraa Abu Elsoud	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Kamal Salah	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	

C	: **C**onceptualization	I	: **I**nvestigation
M	: **M**ethodology	R	: **R**esources
So	: **So**ftware	D	: **D**ata Curation
Va	: **Va**lidation	O	: Writing - **O**riginal Draft
Fo	: **Fo**rmal analysis	E	: Writing - Review & **E**diting

Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [ZS], upon reasonable request.

REFERENCES

[1]   H. Baofang, H. Wang, L. Wang, and W. Yuan. "Adverse drug reaction predictions using stacking deep heterogeneous information network embedding approach," *Molecules*, vol. 23, no. 12, 2018, doi: 10.3390/molecules23123193.
[2]   X. Zhao, L. Chen, Z. H. Guo, and T. Liu. "Predicting drug side effects with compact integration of heterogeneous networks," *Current Bioinformatics*, vol. 14, no. 8, pp. 709-720, 2019, doi: 10.2174/1574893614666190220114644.
[3]   H. Baofang, H. Wang, and Z. Yu. "Drug side-effect prediction via random walk on the signed heterogeneous drug network," *Molecules*, vol. 24, no. 20, 2019, doi: 10.3390/molecules24203668.
[4]   C. C. Yang and M. Zhao. "Mining heterogeneous network for drug repositioning using phenotypic information extracted from social media and pharmaceutical databases," *Artificial Intelligence in Medicine,* vol. 96, pp. 80-92, 2019, doi: 10.1016/j.artmed.2019.03.003.
[5]   L. Peng, C. Yang, Y. Chen, and W. Liu. "Predicting CircRNA-Disease associations via feature convolution learning with heterogeneous graph attention network," *IEEE Journal of BHI* 27, no. 6, pp. 3072-3082, 2023, doi: 10.1109/JBHI.2023.3260863.
[6]   P. Chandak, K. Huang, and M. Zitnik. "Building a knowledge graph to enable precision medicine," *Scientific Data* 10, no. 1, 2023, doi: 10.1038/s41597-023-01960-3.
[7]   G. Khekare, S. Ghugare, R. Khatri, G. Majumder, and U. Khekare, "Blockchain powered integrated health profile and record management system for seamless consultation leveraging unique identifiers," *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, Vellore, India, pp. 1-9, 2024, doi: 10.1109/ic-ETITE58242.2024.10493266.
[8]   Z. Salah, E. Elsoud, K. Salah, W. T. Al-Sit, M. Maaya'a, and A. Al Khawaldeh, "Patient-patient interactions visualization for drug side effects in patients' reviews," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 2007-2020, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp2007-2020.
[9]   A. Bermingham, M. Conway, L. McInerney, N. O'Hare, and A. F. Smeaton. "Combining social network analysis and sentiment analysis to explore the potential for online radicalisation," *In Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM '09*, pp. 231–236, Washington, DC, USA, 2009, doi: 10.1109/ASONAM.2009.31.
[10] P. A. Gloor, J. Krauss, S. Nann, K. Fischbach, and D. Schoder, "Web science 2.0: Identifying trends through semantic social network analysis," *2009 international conference on CSE*, vol. 4, IEEE, 2009, doi: 10.1109/CSE.2009.186.
[11] J. Rabelo, R. B. C. Prudencio, and F. Barros, "Collective classification for sentiment analysis in social networks," *In Proceedings of the 2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, Washington, USA, 2012, pp. 958-963, doi: 10.1109/ICTAI.2012.135.
[12] C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang. "Sentiview: sentiment analysis and visualization for internet popular topics," *IEEE transactions on human-machine systems*, vol. 43, no. 6, pp. 620-630, 2013, doi: 10.1109/THMS.2013.2285047.
[13] M. Shams, M. Saffar, A. Shakery, and H. Faili. "Applying sentiment and social network analysis in user modelling," *In Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing,* Berlin, Heidelberg, pp. 526-539, 2012, doi: 10.1007/978-3-642-28604-9 43.
[14] W. Deitrick and W. Hu. "Mutually enhancing community detection and sentiment analysis on twitter networks," *Journal of Data Analysis and Information Processing*, vol. 1, no. 3, pp. 19-29, 2013, doi: 10.4236/jdaip.2013.13004.
[15] H. Deng, J. Han, H. Ji, H. Li, Y. Lu, and H. Wang. "Exploring and inferring user-user pseudo-friendship for sentiment analysis with heterogeneous networks," *Statistical Analysis and Data Mining: The ASA Data Science Journal,* vol. 7, no. 4, pp. 308-321, 2014, doi: 10.1002/sam.11223.

[16] M. Miller, C. Sathi, D. Wiesenthal, J. Leskovec, and C. Potts, "Sentiment flow through hyperlink networks," *In Proceedings of the International AAAI Conference on WSM*, 2011, vol. 5, no. 1, pp. 550-553, doi: 10.1609/icwsm.v5i1.14199.

[17] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," *In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11*, New York, USA, 2011, pp. 1397-1405, doi: 10.1145/2020408.2020614.

[18] M. Newman. *Networks*. Oxford, United Kingdom: Oxford University Press, 2018.

[19] N. V. Canudas, M. C. Gómez, X. Vilasís-Cardona, and E.G. Ribé, "Graph clustering: a graph-based clustering algorithm for the electromagnetic calorimeter in LHCb," *European Physical Journal C,* vol. 83, no. 2, Feb. 2023, pp. 179, doi: 10.1140/epjc/s10052-023-11332-1.

[20] M. Vernet, Y. Pigné, and E. Sanlaville, "A study of connectivity on dynamic graphs: computing persistent connected components," *4OR: A Quarterly Journal of Operations Research*, vol. 21, no. 2, pp. 205-233, doi: 10.1007/s10288-022-00507-3.

[21] B. Bollobás and O. M. Riordan, "Mathematical results on scale-free random graphs," *Handbook of Graphs and Networks: From the Genome to the Internet,* pp. 1-34, 2003, doi: 10.1002/3527602755.ch1.

[22] M. E. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003, doi: 10.1137/S003361445034248.

[23] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, no. 2, pp. 155-163, 2009, doi: 10.1016/j.socnet.2009.02.002.

[24] J. Mohamadichamgavi, M. Hajihashemi, and K.A. Samani, "An analysis of correlation and comparisons between centrality measures in network models," *Journal of Social Structure,* vol. 25, no. 1, pp. 1-21, 2024, doi: 0-21307/joss-2024-001.

[25] M. Latapy, "Main-memory triangle computations for very large (sparse (power-law)) graphs," *Theoretical Computer Science*, vol. 407, no. 1-3, pp. 458-473, 2008, doi: org/10.1016/j.tcs.2008.07.017.

[26] U. Brandes, "A faster algorithm for betweenness centrality*," Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163-177, 2001, doi: 10.1080/0022250X.2001.9990249.

[27] Q. Wang, N. Xiang, M. You, and X. Rao, "Betweenness centrality approximation in large networks using shortest paths approximation and adaptive sampling," In *International Conference on Internet of Things and Machine Learning (IoTML 2023)*, pp. 278-284, 2023, doi: 10.1117/12.3013414.

BIOGRAPHIES OF AUTHORS
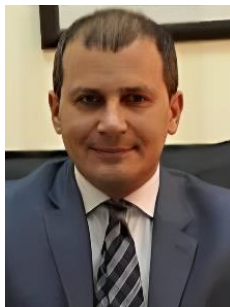
Zaher Salah received his Ph.D. degree in computer science from the University of Liverpool, UK, in 2014, his M.Sc. degree in computer science from Yarmouk University, Jordan, in 2004, and his B.Sc. degree in computer science from University of Jordan, Jordan, in 2001. He is currently an Associate Professor in the Department of Information Technology at The Hashemite University, Zarqa, Jordan. His research interests include machine learning, cyber security, information retrieval, opinion mining, sentiment analysis, biometrics, digital image and analysis, and pattern recognition. He can be contacted at email: zaher@hu.edu.jo.

Esraa Abu Elsoud received the B.Sc. degree in electrical engineering from Hashemite University, Jordan, in 2013 and M.Sc. in cyber security from The Hashemite University in 2023. Her current research interests include cyber security, machine learning, big data, and mobile network. She is currently a lecturer in Zarqa University, Zarqa, Jordan. She can be contacted at email: eabuelsoud@zu.edu.jo.

Kamal Salah received the B.Sc. degree in physics from Yarmouk University, Jordan in 2003, his M.Sc. degree in applied physics (experimental atomic and molecular physics) from The Hashemite University Jordan in 2007. He is currently a lecturer in the Deanship of Preparatory Year and Supporting Studies of the Imam Abdulrahman Bin Faisal University, P.O Box 1982, Dammam, Saudi Arabia. He can be contacted at email: kisalah@iau.edu.sa.