Mathematics Interdisciplinary Research **7** (2022) 179 − 195

Original Scientific Paper

ULPD: An Unsupervised Learning Model to Identify Party and Date Hubs

Mahnaz Habibi ⋆ *and Mitra Habibi*

Abstract

It has been claimed that Protein-Protein Interaction (PPI) networks are scale free that contain a few hubs with ability to bind multiple proteins. Hubs are classified as party and date hubs. Party hubs generally bind different proteins in specific module simultaneously, while date hubs interact with multiple proteins in different modules at different times and locations. Generally, they have been divided into two classes based on the average Pearson Correlation Coefficient (avPCC) of expression over all partners or their functions. In this study, we propose a two-step algorithm to classify party and date hubs based on their topological features of PPI network. In the first step, we calculate some topological features for each hub vertex in PPI network. In the second step, we apply an unsupervised learning model to calculate Laplacian score for each feature. The Laplace value for each hub vertex is considered based on Laplacian scores. Finally, the hub vertices are classified into two classes date hubs and party hubs with respect to Laplace values. We evaluate our method on reference hubs based on the avPCC on PPI network. We show that the combination of topological features based on ULPD can improve the performance of each topological feature. Finally, we investigate the performance of our method for human dataset and analyze two types of hubs as drug targets for Covid-19.

Keywords: party hub, date hub, topological features, unsupervised learning model.

2020 Mathematics Subject Classification: 92B05.

How to cite this article
M. Habibi and M. Habibi, ULPD: An unsupervised learning model to identify party and date hubs, *Math. Interdisc. Res.* **7** (2022) 179 − 195.

⋆Corresponding author (E-mail: m_habibi@qiau.ac.ir)
Academic Editor: Hikmet Koyunbakan
Received 10 September 2021, Accepted 17 April 2022
DOI: 10.22052/mir.2022.243082.1308

1. Introduction

Proteins are identified as the main agent of biological processes that can determine the phenotype of organisms. Some proteins are functional isolated form and some ones interact with other proteins or other molecules. These interactions between the proteins are often represented in the form of Protein-Protein Interaction (PPI) network [3, 4, 9, 21, 24]. It is desirable to analyze these networks to understand the internal organization of the cells. Since some proteins interact with multiple proteins and others interact with only a few, the PPI network has a wide range of degrees. The highly connected proteins in PPI network are referred as hubs. There are some studies that reveal the functional and structural characterization of hubs in PPI network [1, 8, 11, 19]. Bertolazzi et al. have determined the special biological and structural features of hub proteins. They have studied the role of hubs in molecular organization of PPI networks [8]. In 2010, Kukuroghu et al., have examined the flexibility of the hubs using structural perspective [11]. Recently some computational methods have been presented the implication of targeting hub proteins in treatment of diseases [2, 26, 28]. In the recent pandemic, Covid-19 (coronavirus disease), the role of hub proteins in virus-host interaction network have been highlighted to study pathogenesis of infection [26]. Prased et al. have analyzed the human PPI network and targeted hub proteins to find candidate drugs for Covid-19 [28].

In 2004, Han et al. have expressed that hub proteins which interact with most of their partners simultaneously are designated as party hubs and those that bind their different partners at different times or locations are date hubs [16]. Briefly, they have divided the highly connected proteins into two classes based on the average of Pearson Correlation Coefficient (avPCC) of expression over all partners. They have studied the role of two types of hubs in molecular organization in a cell. They have also shown that party and date hubs play an important role in organizing modules. Some topological features of two types of hub proteins in the PPI network are presented [7, 12, 14, 27, 34]. Centrality measures are the one of the main topological features that studied by Yu et al. [34]. They have defined bottlenecks as proteins with a high betweenness centrality. They have believed that date hubs are bottlenecks with the high degree while the party hubs correspond mostly to hubs with the low betweenness in the PPI network. Other topological features of party and date hubs in the PPI network are considered by Bertin et al. [7]. They have shown that removing date and party hubs shows different behavior on PPI network. In the other word, removing hubs from PPI network have distinct effect on the characteristic path length.

In this study, we present an Unsupervised Learning model to identify Party and Date hubs (ULPD). ULPD is a computational method based on a combination of topological features of PPI network to classify party and date hubs. In the first step, we propose informative features based on centrality mature for each vertex on PPI network. In the second step, Laplacian score is calculated for each feature based on Laplacian Eigenmaps [6] and Locality Preserving Projection [17]. Finally,

hub vertices are classified based on Laplace values that calculated for hub vertices. Then, we evaluate each topological feature and their combination with respect to precision, recall and F-measure. Results show that the combination of these features performs better than each existing feature on PPI network. Finally, we analyze party and date hubs for human PPI netwok. Our evaluation show that date hubs are some appropriate candidate drug targets. Most of these candidate targets are recommended in other studies.

2. Methods

In a topological sense, a PPI network is considered as undirected graph $G =< V, E >$. Generally, a vertex v of PPI network represents a given protein and each edge uv between two vertices u and v represents the functional interaction between two proteins. A vertex u is neighbor of another vertex v, if uv is an edge of G. The set of all vertices that are the neighbors of v is the neighborhood of v and it is denoted by $N(v)$. The number of vertices of $N(v)$ is called the degree of v and denoted by $d(v)$. Let H be a subset of vertices of G, the neighborhood of H is the union of the neighborhood of each vertex of H, without H. It is defined as following formula:

$$N(H) = \cup_{v \in H} N(v) - H\,.$$

A sequence $u = u_0, u_1, \ldots, u_n = v$ of distinct vertices of non-empty network G is a path between two vertices u and v, if $u_i u_{i+1}$ for each $0 \leq i < n$ is an edge of G. The number of edges of a path is considered as the length of the path. The shortest path between two vertices u and v is defined as a path with the minimum length. The minimum length of path between two vertices is denoted by $d(u, v)$. The following eight features show the informative topological features for each vertex in PPI network.

The Clustering Coefficient (CC) for each vertex v of $G = \langle V, E \rangle$ is defended as following formula:

$$CC(v) = \frac{2|E(H)|}{d(v)(d(v)-1)},$$

where $H = N(v)$ and $|E(H)|$ is the number of $\{uw \in E; u, w \in N(v)\}$. If all vertices of H are pair wise neighbor, then H is a complete subgraph of G and $CC = 1$. Generally, the clustering coefficient denotes the density of the neighbors of each vertex.

The Betweenness (Bn) centrality for each vertex v of $G = \langle V, E \rangle$ is defined by:

$$Bn(v) = \sum_{u,w \in V} \frac{\tau_{uw}(v)}{\tau_{uw}},$$

where τ_{uw} is the total number of shortest paths from vertex u to vertex w for each u and w of V and $\tau_{uw}(v)$ is the number of these shortest paths that pass through v. It denotes the control of each vertex over the flow of information in the network.

Let $S = v \cup N(v)$, the Expansion (Ex) of v is defined by:

$$Ex(v) = \frac{|N(S)|}{|S|},$$

where $|N(S)|$ and $|S|$ are the size of $N(S)$ and S respectively.

Also the Density (D) of v is defined as following formula:

$$D(v) = \frac{2|E(S)|}{|S|(|S|-1)},$$

where $S = v \cup N(v)$ and $|E(S)| = |\{uv \in E; u, v \in S\}|$.

Suppose $C = \{c_1, c_2, \ldots, c_n\}$ is the clusters of vertices of graph $G = \langle V, E\rangle$. The Participation Coefficient (PC) of each vertex v based on given cluster C is calculated by:

$$PC(v) = 1 - \sum_{i=1}^{n} (\frac{d_i(v)}{d(v)})^2,$$

where $d_i(v)$, $1 \leq i \leq n$, is the number of neighbors of v within c_i. The participation coefficient value for each vertex is a number between 0 and 1. If $PC(v)$ approaches 1, then we say the vertex v participate in maximum number of clusters. To calculate participation coefficient value for each vertex on PPI network, we use the Markov Cluster algorithm (MCL) [32].

The Closeness centrality (Cl) measure for each vertex v of network $G = \langle V, E\rangle$ is defined by:

$$Cl(v) = \frac{|V|-1}{\sum_{u \in V} d(u, v)},$$

If the $Cl(v)$ meets or exceeds a threshold, then we say the vertex v has the shortest distance with the other vertices of network.

Another centrality measure of each vertex v on graph $G = \langle V, E\rangle$ is Subgraph Centrality (SC). The subgraph centrality for each vertex is defined as following formula:

$$SC(v) = \sum_{k=1}^{\infty} \delta_k(v),$$

where $\delta_k(v)$ is the number of path with length k that pass through v.

Finally, The Mean Degree Neighbor (MDN) for each vertex is calculated as a topological feature of G. Briefly, it is calculated as following formula:

$$MDN(v) = \frac{\sum_{u \in N(v)} d(u)}{|N(v)|}.$$

In this work, we propose a new algorithm named ULPD (Unsupervised Learning model to identify Party and Date hubs) from input data. The ULPD algorithm comprises two main steps. In the first step, some mentioned topological features

are calculated for each hub vertex in PPI network. In the second step, we propose an unsupervised learning approach based on Laplacian Eigenmaps [6] and Locality Preserving Projection [17]. In following section, we describe the details of Laplacian Score as a feature selecting method which are mapped from algorithm reported by He et al. [18].

Let $A = [a_{ij}]$ shows the feature matrix of PPI network G with n hub vertices such that a_{ij} denote the j-th feature of i-th hub vertex ($1 \leq j \leq 8$). Let $A_j = [a_{1j}, \ldots, a_{nj}]^T$ be the column matrix corresponding j-th feature of hub vertices and let $\vec{F_i} =< a_{i1}, \ldots, a_{i8} >$ be the feature vector of each hub vertex v_i. The weighted matrix $S = [s_{ij}]_{n \times n}$ is defined as follow:

$$s_{ij} = \begin{cases} e^{-\frac{|\vec{F_i} - \vec{F_j}|^2}{2}} & if\ |\vec{F_i} - \vec{F_j}| < \delta, \\ 0 & otherwise. \end{cases}$$

We define a matrix $D = diag(d_i)$ as a diagonal weighted matrix S, where $d_i = \sum_{k=1}^{n} s_{ik}$.

Now the Laplacian matrix of S is considered as $L = S - D$. For j-th feature, the column matrix is calculated by:

$$\tilde{A_j} = A_j - \frac{A_j^T DJ}{J^T DJ},$$

where $J = [1\ 1\ \ldots\ 1]^T$. Now, Laplacian Score for j-th feature is calculated by follow:

$$\tilde{L_j} = \frac{\tilde{A_j}^T L \tilde{A_j}}{\tilde{A_j}^T D \tilde{A_j}}.$$

Finally, for i-th hub vertex of network G the Laplace value is calculated as following formula:

$$L(v_i) = \sum_{j=1}^{8} a_{ij} \rho_j L_j,$$

where $\rho_j = -1$ for three topological features (Participation Coefficient, Closeness and Betweenness) and for other features $\rho_j = 1$. The set of hub vertices are classified into two clusters. The vertices that meet or exceed a threshold θ are considered as date hubs and other vertices are considered as party hubs.

3. Results

3.1 Datasets

In this work, we use two high-throughput protein-protein interaction data collections of Saccharomyces cerevisiae [13]. The first dataset collection contains yeast

Table 1: The summery of each dataset: In the datasets column refers to networks, where (1) denotes the first dataset, (2) denotes the second dataset. Here |V| and |E| are the number of genes and interactions between them respectively and |Hub| is the number of hub proteins. The number of refrence party and date hub proteins are denoted by |Party| and |Date| respectively.

Datasets	|V|	|E|	|Hub|	avPCC	|Party|	|Date|
(1)	4445	45869	465	0.65	127	338
(2)	2834	41258	290	0.58	94	196

cells which grown aerobically on galactose medium. Another data collection is industrial fermentations of saccharomyces cerevisiae which grown in glucose-limited chemostat culture in different oxygen concentrations [16]. CLR algorithm [7] identifies 45869 regular interactions between 4445 genes in the first dataset and the second dataset contains 2834 genes and 41258 interactions. We consider ten percentages of high-degree nodes as hubs. This yields 465 hubs in the first dataset and 290 hubs in the second dataset. We identify real date and party hubs according to the definition of Han et al. [16]. We calculate the average of Pearson Correlation Coefficients (avPCC) between the hub and each of its respective partners for mRNA expression to distinct date and party hubs. Ones with relatively high avPCCs are chosen as real party hubs and the other ones are defined as real date hubs. Some descriptive statistics of each protein interaction network are presented in Table 1.

3.2 Performance Evaluation Measures

To evaluate the performance of our method, we use some evaluation measures. These measures are based on the relation between the number of hubs correctly predicted positive (TP), the number of hubs correctly predicted negative (TN), the number of hubs incorrectly predicted positive (FP), and the number of hubs incorrectly predicted negative (FN). The precision (Pre) and recall (Re) are two evaluation measures and they are defined as following formula:

$$Pre = \frac{TP}{Tp + Fp},$$

$$Re = \frac{TP}{Tp + Fn}.$$

They are two numbers between 0 and 1 and commonly used to evaluate the performance of classification methods. In particular, precision corresponds to the fraction of prediction hub class that are matched by real hub class and recall corresponds to the fraction of real hub class that are matched by predicted hub class.

Table 2: The threshold values of each criterion with respect to the best F-measure.

Topological features	Thresholds
Subgraph Centrality	0.055
Participation Coefficient	0.85
Expansion	0.22
Density	0.59
Mean Degree Neighbor	0.54
Clustering Coefficient	0.6
Closeness centrality	0.5
Betweenness	0.08

Another measure which can be used to evaluate the performance of a method is F-measure. The F-measure as the harmonic mean of precision and recall is defined as follow:

$$F = \frac{2\ Precision\ Recall}{Precision + Recall}.$$

3.3 Testing for Accuracy

The previous studies on party and date hubs show that the centrality measures are the main topological features to classify hubs [34]. In this work, we used some centrality measures as topological features, such as clustering coefficient, betweenness, participation coefficient, closeness, subgraph centrality, mean degree neighbor, density and expansion to identify party and date hubs. The histogram of frequency distribution of each normalized topological feature for real date and party hubs on the first dataset is shown in Figure 1. In this figure, the blue lines indicate the date hubs and the red lines indicate the party hubs. The figure shows that there are no specific threshold on each topological feature to classify party and date hubs completely.

To check the validity of ULPD, we compare the output of ULPD with the two classes of hubs obtained by each topological feature, on the two protein interaction networks. The party and date hubs which obtained from avPCC definition are used as benchmark real party and date hubs. The best threshold value of each topological feature is calculated on the first datset. By varying the threshold on each normalized topological feature, we calculate the harmonic mean of F-measure values based on two classes of hubs. The best threshold value for each topological feature is presented in Table 2. In this survey, the topological features are modeled by Laplacian matrix. The output of this model is Laplacian Score for each feature. All Laplacian Scores which calculated by our model are shown in Table 3. We also find that the best threshold of Laplace value to classify date and party hubs on first dataset are 0.3.

In the first dataset, almost all features have similar performance in recall how-

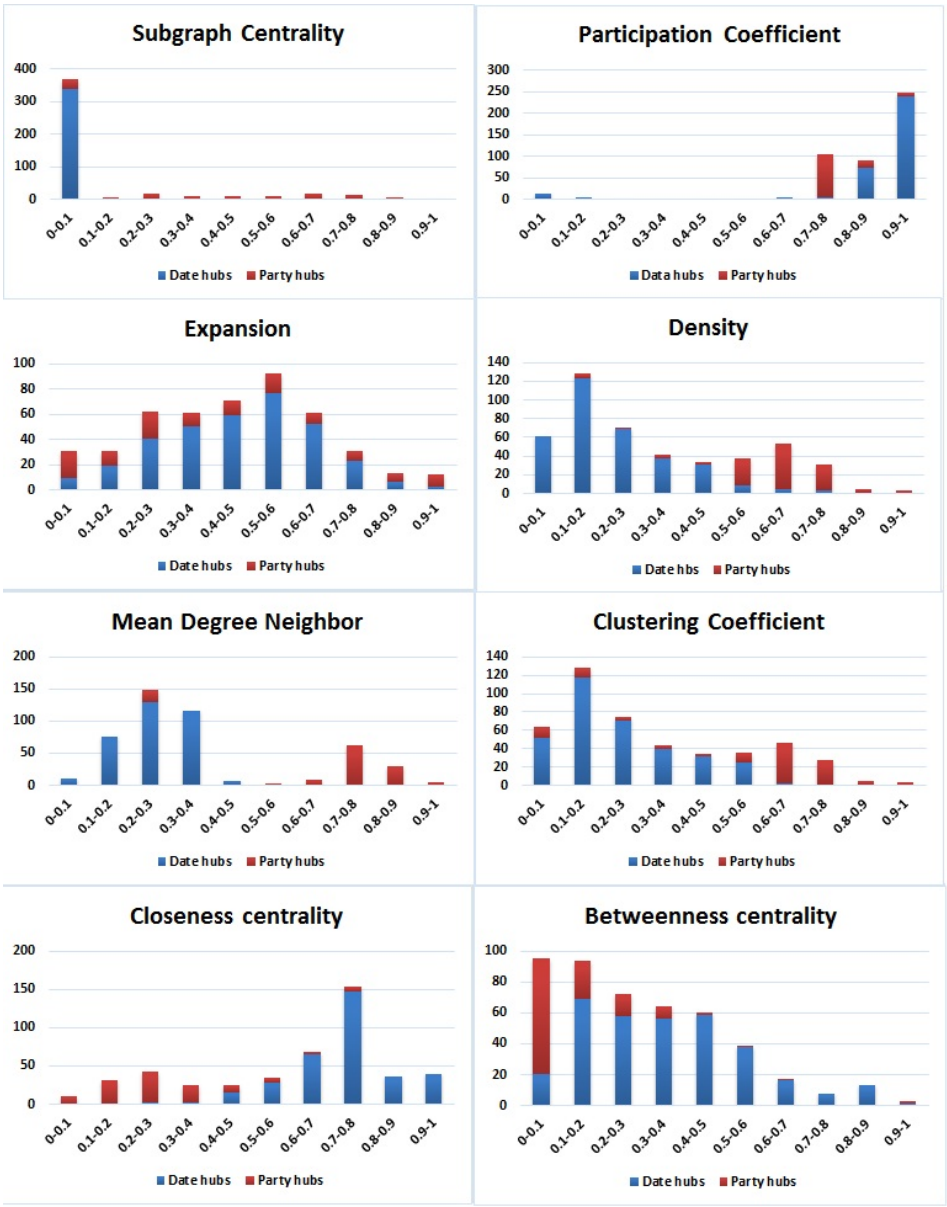

Figure 1: the histogram of frequency distribution of each topological feature for date and party hubs on first dataset.

Table 3: The Laplacian Score of each criterion which is modeled by unsupervised learning model.

Topological features	Laplacian Scores
Subgraph Centrality	0.97
Participation Coefficient	0.95
Expansion	0.12
Density	0.74
Mean Degree Neighbor	0.79
Clustering Coefficient	0.65
Closeness centrality	0.71
Betweenness	0.43

Table 4: Precision and recall values of ULPD and other features on first dataset to predict date hubs.

	TP	TN	FP	FN	Pre	Re	F
ULPD	337	117	10	1	**0.97**	**0.99**	**0.984**
Subgraph Centrality	337	107	20	10	0.94	**0.99**	0.96
Participation Coefficient	302	104	23	36	0.92	0.89	0.91
Expansion	303	72	55	35	0.84	0.89	0.87
Density	325	110	17	13	0.95	0.96	0.95
Mean Degree Neighbor	337	105	23	1	0.90	**0.99**	0.96
Clustering Coefficient	325	110	17	13	0.95	0.96	0.95
Closeness	328	113	14	10	0.95	0.97	0.96
Betweenness	325	69	58	13	0.84	0.96	0.90

ever, as can be seen in Table 4, ULPD performs better than others to predict date hubs. Furthermore, this significant superiority of ULPD in recall comes with the highest precision value. Also, Table 5 show the superiority of ULPD in both precision and recall values to predict party hubs on the first datset.

The first dataset is a difficult example because, the best threshold values are chosen with respect to maximum F-measures on the first dataset. Then, we compare ULPD results with the results of topological features on the second dataset. We find that ULPD shows the best performance compare to other features to predict party and date hubs on the second dataset. In fact, both precision and recall values of ULPD are greater than all features, as can be seen in Table 6 and Table 7.

Table 5: Precision and recall values of ULPD and other features on first dataset to predict party hubs.

	TP	TN	FP	FN	Pre	Re	F
ULPD	117	337	1	10	**0.99**	**0.92**	**0.95**
Subgraph Centrality	107	337	10	20	**0.99**	0.84	0.91
Participation Coefficient	104	302	36	23	0.74	0.81	0.77
Expansion	72	303	35	55	0.67	0.56	0.61
Density	110	325	13	17	0.89	0.86	0.89
Mean Degree Neighbor	105	337	1	23	**0.99**	0.82	0.89
Clustering Coefficient	110	325	13	17	0.89	0.86	0.88
Closeness	113	328	10	14	0.91	0.88	0.90
Betweenness	69	325	13	58	0.86	0.56	0.66

Table 6: Precision and recall values of ULPD and other features on second dataset to predict date hubs.

	TP	TN	FP	FN	Pre	Re	F
ULPD	186	82	12	10	**0.93**	**0.94**	**0.94**
Subgraph Centrality	175	77	17	21	0.91	0.89	0.90
Participation Coefficient	183	75	19	13	0.90	0.93	0.91
Expansion	176	33	61	20	0.74	0.89	0.81
Density	177	74	20	19	0.89	0.90	0.90
Mean Degree Neighbor	179	70	24	17	0.88	0.91	0.89
Clustering Coefficient	179	75	19	17	0.90	0.91	0.90
Closeness	176	70	24	20	0.88	0.89	0.88
Betweenness	182	51	43	14	0.92	0.80	0.86

Table 7: Precision and recall values of ULPD and other features on second dataset to predict party hubs.

	TP	TN	FP	FN	Pre	Re	F
ULPD	82	186	10	12	**0.89**	**0.87**	**0.88**
Subgraph Centrality	77	175	21	17	0.78	0.81	0.80
Participation Coefficient	75	183	13	19	0.85	0.79	0.82
Expansion	33	176	20	61	0.62	0.35	0.44
Density	74	177	19	20	0.79	0.78	0.79
Mean Degree Neighbor	70	179	17	24	0.80	0.74	0.77
Clustering Coefficient	75	179	17	19	0.81	0.79	0.80
Closeness	70	176	20	24	0.77	0.74	0.76
Betweenness	51	182	14	43	0.78	0.54	0.64

Table 8: Precision and recall values of ULPD and other features on second dataset to predict party hubs.

	Gene Name
Green-Date	ACD, ACTB, ATM, B2M, BRCA1, BTK, CARD9, CASP3, CASP8, CCL5, CD40, CD79A, CD81, CDC42, CFTR, DICER1, EGFR, ERBIN, FADD, FANCA, FANCC, FANCG, FAS, FASLG, FOXP3, GATA1, HAX1, HDAC6, HTRA2, HYOU1, IGHM, IKBKB, IKZF1, IRAK1, IRF3, ISG15, ITCH, ITGB3, JAK1, JAK3, LAT, LCK, LIG4, MSN, MYD88, NCSTN, NFE2L2, NFKB1, NFKBIA, PIK3R1, PLCG2, POLD1, POLR3A, POMP, PRKCD, PRKDC, PSEN1, PSMB8, PSTPIP1, PTEN, RAD51, RBCK1, REL, RELA, SEC61A1, SH3KBP1, SLX4, STAT1, STAT3, TAZ, TBK1, TCF3, TFRC, TGFB1, TGFBR1, TGFBR2, TINF2, TNF, TNFRSF1A, TP53, TRAF3, TYK2, WAS, XIAP, ZAP70, IKBKG
Green-Par ty	ADAR, AIRE, ARPC1B, BCL10, BLM, BRCA2, COPA, CTPS1, DKC1, DNMT3B, FANCD2, FANCI, G6PD, IL7R, KMT2A, MAP3K14, MCM4, MSH6, MTHFD1, NBN, NFKB2, NOS2, PMS2, RANBP2, RELB, RIPK1, RNF31, RPSA, SAMHD1, SMARCD2, STAT5B, STK4, STX11, TERT, TNFAIP3, TOP2B, WDR1

3.4 Performance of our Method for Human PPI Network

The previous studies on hub proteins show that they could be particularly interesting drug targets [5]. In this work, we study party and date hub proteins as drug targets for COVID-19. We propose a human PPI network that obtained by habibi et al. [15]. This PPI network is gathered by five human high-throughput binary interactions [3, 4, 9, 21, 24]. Then, some proteins which cannot be mapped to a Uniprot ID [31] are removed. It yields to 304730 interactions between 25260 proteins. We choose ten percentages of high-degree vertices in human PPI network as hubs. Applying ULPD algorithm on PPI network reveals 704 party and 1824 date hubs among 2528 human hubs.

We also use a research gene panel associated with SARS-CoV-2 to identify essential genes related to disease. The high level of evidence genes are gathered by [25] (COVID-19 research (Version 1.79)). It contains 461 genes related to SARS-CoV-2. We find that from 461 genes 86 genes have been selected as date hubs that denoted as Green-Date hubs and 37 genes have been identified as party hubs that denoted as Green-Party hubs. Table 8 shows these Green-Date and Green-Party genes that are related to SARS-CoV-2.

To identify the subset of date hubs and party hubs as suitable candidate genes with important biological roles related to COVID-19, we narrow down Green-Party and Green-Date hubs to the disease-associated genes. The previous studies show that various symptoms from asymptomatic to death between different patient with

COVID-19. These studies indicated that patient with underlying diseases such as, Cardiovascular diseases, Diabetes, Hepatitis, Lung diseases, Kidney disease, and different types of cancers have more severe symptoms than others. Then we correlate the genes related to these diseases in in Green-Date and Green-party sets. Then we find three genes, ITGB3, TNF, TP53 in Green-Date set are corresponding to five out of six of mentioned diseases.

The first main gene is ITG-B3. Integrins (ITBs) constitute an important sub-family and have a meaningful correlation with enhanced invasion. It is noticeable that the high proportions of these genes were detected in the lung of the COVID-19 patients [20]. Researchers observed that this high level of ITBs might predispose to greater on cell entry. The study on this gene could be a promising therapeutics target [30].

The second main gene is TNF that is considered as cytosine. Tumor Necrosis Factors (TNFs) are a super-family that by used by immune system cells. They are related to proinflammatory responses and play main role in the regulation a inflammatory processes. In the most severe cases, the high level of TNF alpha is observed [10, 22]. Leng et al., observed that intravenous injection of MSCs have been significantly decreased the pro-inflammatory cytosine TNF-α and it can be possible therapeutic approach to inhibit TNF-α, among other cytokines [23].

The next main hub gene related to COVID-19 is TP53 (P53). Cellular tumor antigen p53 is a key participant in the type I interferon antiviral defense mechanism [23]. The spike protein of virus using p53 degradation tries to help viral survival in lung cells. Researcher believes that p53 might effectively inhibit viral replication in the human respiratory tract and lung cells [29].

To evaluate party and date hubs associated with SARS-CoV-2, we also study 449 drugs as clinical trials that are reported in Drug Bank [33]. We consider some drugs if they have two features. First, it needs to be studies in more than two clinics. Second, it needs to have human targets in PPI network. Finally, we select 328 drugs that denoted by Clinical-Drug. Evaluation of protein targets in Clinical-Drug group shows that from 328 drugs in this group, none of them is approved by Green-Party hubs. But 21 drugs of Clinical-Drug group are approved by Green-Date hubs. We also find that from 86 genes in Green-Date set, 18 genes are targeted by this group of drugs.

4. Summery and Discussion

One of the beneficial filed of research is the analysis of protein-protein interaction network and its importance has been increasing recently. For example, it helps in discovering unknown essential proteins related to diseases. Hub proteins as highly connected proteins in PPI network give more information on essential proteins. The main idea of this work is to classify hub proteins into two classes (date and party hubs) based on topological features, and evaluate two hub classes as drug targets on human PPI network.

In the first part of this work, we have presented the short preliminaries of graph theory and mentioned some topological features of each hub vertex in PPI network. Then, we have proposed the ULPD algorithm to identify party and date hubs based on unsupervised learning model.

In the second part of this work, we have studied the impact of each topological feature to identify party and date hubs. The results on two datasets show that our algorithm can progressively improve the performance of each centrality feature to identify party and date hubs based on common evaluating parameters (TP, TN, FP, FN and F-measure). Results indicate that ULPD algorithm agrees well with two hub classes obtained by the average Pearson Correlation Coefficient between hub and each of respective partners for mRNA expression, and F-measure values can be increased considerably in comparison with each feature.

Finally, we have performed our method to identify party and date hubs for human PPI network. We have studied the party and date hubs as two different groups of essential proteins. In the first part, 1824 date hubs and 704 party hubs for human PPI network are identified. Since it seems urgent to find the essential genes associated with COVID-19 in the recent time, we have focused on identifying the hubs that are essential to the pathology of this disease. We found a subset of date hubs and a subset of party hubs that in the previous studies are reported as high-evidence genes related to COVI-19. Finally, we evaluate three essential date hub genes (**ITG-B3**, **TNF**, **TP53**) associated with at least five of six underlying diseases. We find that 8 drugs in Clinical-Drug group including (**Adalimumab**, **Chloroquine**, **Infliximab**, **Pomalidomide**, **Tirofiban**, **Pomalidomide**, **Thalidomide**, **Aspirin**) are approved by these three hub genes. Our study showed that date hub proteins can be effective and suitable candidates in clinical trials for COVID-19 treatment.

Conflicts of Interest. The authors declare that there are no conflicts of interest regarding the publication of this article.

References

[1] S. Agarwal, C. M. Deane, M. Porter and N. S. Jones, Revisiting date and party Hubs: Novel approaches to role assignment in protein interaction networks, *PLoS Comput Biol.* **6** (2010) e1000817.

[2] V. K. MD. Aksam, V. M. Chandrasekaran and S. Pandurangan, Hub nodes in the network of human Mitogen-Activated protein Kinase (MAPK) pathways: Characteristics and potential as drug targets, *Inform. Med. Unlock.* **9** (2017) 173 − 180.

[3] G. Alanis-Lobato, M. A. Andrade-Navarro and M. H. Schaefer, HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks, *Nucleic Acids Res.* **45** (2017) D408−D414.

[4] D. Alonso-Lopez, F. J. Campos-Laborie, A. Miguel, A. errez, L. Lambourne, M. A. Calderwood, M. Vidal and J. D. L. Rivas, APID database: redefining protein-protein interaction experimental evidences and binary interactomes, *Database* **2019** (2019) 1 − 8.

[5] G. Apic, T. Ignjatovic, S. Boyer and R. B. Russell, Illuminating drug discovery with biological pathways, *FEBS Lett.* **579** (2005) 1872 − 1877.

[6] M. Belkin and P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, *Adv. Neural Inf. Process. Syst.* **14** (2001) 585−591.

[7] N. Bertin, N. Simonis, D. Dupuy, M. E. Cusick, J. J. Han, H. B. Fraser, F. P. Roth and M. Vidal, Confirmation of organized modularity in the yeast interactome, *PLoS Biol.* **5** (2007) e153.

[8] P. Bertolazzi, M. E. Bock and C. Guerra, On the functional and structural characterization of hubs in protein-protein interaction networks, *Biotechnol. Adv.* **31** (2013) 274 − 286.

[9] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N. k. Kolas, L. Donnell, S. Oster, C. Theesfeld, A. Sellam, C. Stark, B. J. Breitkreutz, K. Dolinski and M. Tyers, The BioGRID interaction database: 2017 update, *Nucleic Acids Res.* **45** (2016) D369−D379.

[10] V. J. Costela-Ruiz, R. Illescas-Montes, J. M. Puerta-Puerta, C. Ruiz and L. Melguizo-Rodríguez, SARS-CoV-2 infection: The role of cytokines in COVID-19 disease, *Cytokine and Growth Factor Rev.* **54** (2020) 62–75.

[11] E. Cukuroglu, E. Ozkirimli and O. Keskin, Structural properties of hub proteins, 2010 5*th International Symposium on Health Informatics and Bioinformatics*, Ankara, Turkey, 2010.

[12] D. Ekman, S. Light, K. Bjorklund and A. Elofsson, What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae, *Genome Biol.* **7** (6) (2006) R45.

[13] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, J. Cottarel, S. Kasif, J. C. Collins and T. S. Gardner, Large-Scale Mapping and Validation of Escherichia coli Transcriptional Regulation from a Compendium of Expression Profiles, *PLoS Biol.* **5** (2007) e8.

[14] P. S. Gajinder, G. Mythily and D. Debasis, Role of intrinsic disorder in transient interactions of hub proteins, *Proteins: Stru. Func. Bioinf.* **66** (4) (2006) 761 − 765.

[15] M. Habibi, G. Taheri and R. Aghdam, A SARS-CoV-2 (COVID-19) biological network to find targets for drug repurposing, *Scientific Reports* **11** (2021) 9378.

[16] J. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth and M. Vidal, Evidence for dynamically organized modularity in the yeast protein–protein interaction network, *Nature* **430** (2004) 88 – 93.

[17] X. He and P. Niyogi, Locality preserving projections, *Adv. Neural Inf. Process. Syst.* **16** (1) (2003) 186 – 197.

[18] X. He, C. Deng and N. Partha, Laplacian score for feature selection, *Adv. Neural Inf. Process. Syst.* **18** (2005) 507–514.

[19] G. Hu, Z. Wu, V. Uversky and L. Kurgan, Functional analysis of human hub proteins and their interactors involved in the intrinsic disorder-enriched interactions, *Int. J. Mol. Sci.* **18** (12) (2017) 2761.

[20] A. B. M. M. K. Islam, M. A. A. K. Khan, R. Ahmed, M. S. Hossain, S. M. T. Kabir, M. S. Islam and A. Z. Siddiki, Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2, *J. Transl. Med.* **19** (1) (2021) 32.

[21] D. Jishnu and Y. Haiyuan, HINT: High-quality protein interactomes and their applications in understanding human disease, *BMC Syst. Biol.* **6** (2012) 92.

[22] S. Kany, J. T. Vollrath and B. Relja, Cytokines in inflammatory disease, *Int. J. Mol. Sci.* **20** (2019) 6008.

[23] Z. Leng, R. Zhu, W. Hou, Y. Feng, Y. Yang, Q. Han and R. C. Zhao, Transplantation of ACE2-mesenchymal stem cells improves the outcome of patients with COVID-19 pneumonia, *Aging. Dis.* **11** (2020) 216.

[24] K. Luck, D. K. Kim, L. Lambourne, K. Spirohn, B. E. Begg, W. Bian, R. Brignall, T. Cafarelli, F. J. Campos-Laborie, B. Charloteaux, D. Choi, A. G. Coté, M. Daley, S. Deimling, A. Desbuleux, A. Dricot, M. Gebbia, M. F. Hardy, N. Kishore, J. J. Knapp, I. A. Kovács, I. Lemmens, M. W. Mee, J. C. Mellor, C. Pollis, C. Pons, A. D. Richardson, S. Schlabach, B. Teeking, A. Yadav, M. Babor, D. Balcha, O. Basha, C. Bowman-Colin, S. F. Chin, S. G. Choi, C. Colabella, G. Coppin, C. D'Amata, D. De Ridder, S. De Rouck, M. Duran-Frigola, H. Ennajdaoui, F. Goebels, L. Goehring, A. Gopal, G. Haddad, E. Hatchi, M. Helmy, Y. Jacob, Y. Kassa, S. Landini, R. Li, N. van Lieshout, A. MacWilliams, D. Markey, J. N. Paulson, S. Rangarajan, J. Rasla, A. Rayhan, T. Rolland, A. San-Miguel, Y. Shen, D. Sheykhkarimli, G. M. Sheynkman, E. Simonovsky, M. Taan, A.Tejeda, V. Tropepe, J. C. Twizere, Y. Wang, R. J. Weatheritt, J. Weile, Y. Xia, X. Yang, E. Yeger-Lotem, Q. Zhong, P. Aloy, G. D. Bader, J. De Las Rivas, S. Gaudet, T. Hao, J. Rak, J. Tavernier, D. E. Hill, M. Vidal, F. P. Roth and M. A. Calderwood,

A reference map of the human binary protein interactome, *Nature* **580** (2020) 402 − 408.

[25] A. R. Martin, E. Williams, R. E. Foulger, S. Leigh, L. C. Daugherty, O. Niblock, I. U. S. Leong, K. R. Smith, O. Gerasimenko, E. Haraldsdottir, E. Thomas, R. H. Scott, E. Baple, A. Tucci, H. Brittain, A. de Burca, K. Ibañez, D. Kasperaviciute, D. Smedley, M. Caulfield, A. Rendon and E. M. McDonagh, PPanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels, *Nat Genet.* **51** (11) (2019) 1560–1565.

[26] F. Messina, F. and E. Giombini, C, Agrati, F. Vairo, T. A. Bartoli, S. A. Moghazi, M. Piacentini, F. Locatelli, G. Kobinger, M. Maeurer, A. Zumla, M. R. Capobianchi, F. N. Lauria and G. Ippolito, COVID-19: viral-host interactome analyzed by network based-approach model to study pathogenesis of SARS-CoV-2 infection, *J. Transl. Med.* **18** (2020) 233.

[27] M. Mirzarezaee, B. N. Araabi and M. Sadeghi, Features analysis for identification of date and party hubs in protein interaction network of Saccharomyces Cerevisiae, *BMC Sys. Bio.* **4** (2010) 172.

[28] K. Prasad, F. Khatoon, R. Rashid, N. Ali, A. F. AlAsmari, M. A. Ahmed, A. S. Alqahtani, M. S. Alqahtani and V. Kumar, Targeting hub genes and pathways of innate immune response in COVID-19: A network biology perspective, *Int. J. Biol. Macromol.* **163** (2020) 1 − 8.

[29] M. J. Ramaiah, mTOR inhibition and p53 activation, microRNAs: The possible therapy against pandemic COVID-19, *Gene Rep.* **20** (2020) 100765.

[30] C. J. Sigrist, A. Bridge and P. Le Mercier, A potential role for integrins in host cell entry by SARS-CoV-2, *Antiviral Res.* **177** (2020) 104759.

[31] C. UniProt Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.* **47** (D1) (2019) D506−D515.

[32] S. Van Dongen, *Graph Clustering by Flow Simulation,* Ph. D. Thesis Center for Mathematics and Computer Science (CWI), University of Utrecht, 2000.

[33] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant and M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* **46** (D1) (2018) D1074−D1082.

[34] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov and M. Gerstein, The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics, *PLoS Comput. Biol.* **3** (2007) e59.

Mahnaz Habibi
Department of Mathematics,
Qazvin Branch,
Islamic Azad University,
Qazvin, Iran
e-mail: m_habibi@qiau.ac.ir

Mitra Habibi
M. D. of Fazel Araaghi Hospital,
Tehran, Iran
e-mail: mhabibi@ipm.ir