International Journal of
Analytical, Pharmaceutical and Biomedical Sciences

Volume: 4: Issue-3: March-2015 *ISSN:2278-0246*

www.ijapbs.com

CYTOSCAPE RETRIEVED PROTEIN-PROTEIN INTERACTION (PPI) NETWORKS: SEEKING THE CORRELATION OF HUMAN PROTEINS' TOPOLOGICAL FEATURES BETWEEN MAJOR PUBLIC PPI DATABASES DUE TO THEIR MEDICAL IMPORTANCE

Nahid Safari-Alighiarloo[1], Mohammad Taghizadeh[2]*, Mostafa Rezaei-Tavirani[3]*

[1, 3]Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[2]Bioinformatics Department, Institute of Biochemistry and Biophysics, Tehran University, Tehran, Iran.
*Corresponding authors; Email addresses: 1- Nahid Safari-Alighiarloo1 N.Safari@sbmu.ac.ir
2- Mohammad Taghizadeh2 mtaghizadeh@ut.ac.ir
3- Mostafa Rezaei-Tavirani1 tavirany@yahoo.com telephone: +98-2122714248
fax: +98-2122714248.

ABSTRACT: Background: Protein-protein interaction (PPI) databases have become major sources to study networks and cellular pathways, which can be utilized to get invaluable information in biomedical researches. There are some public PPI databases have already deposited the large amount of experimentally verified interactions from literatures using different curation policies. Our aim was to evaluate to what extent common human proteins' topological features have correlations between these databases. Methods: Five major public PPI databases (DIP, MINT, HPRD, IntAct and BIND) obtained from Cytoscap. Using statistical analysis such as calculation of Correlation coefficients, seven topological features of some common human proteins were compared between these databases. Results: The results showed that some medium and weak significant correlations of shared proteins' topological features were between five databases, and one robust correlation for degree was found between HPRD and BIND (Spearman's rank r = 0.79, P = 0) databases. Conclusions: Several lines of evidence indicated that PPI data have been reported according to different organizing strategies of databases; as a result, this issue may affect the structure of derived PPI networks. Probably because of this reason, our results confirm this point that topological features of proteins in current human PPI networks are highly network dependent.
Keywords: Protein-Protein Interaction, PPI, Databases, Topological Features, System Biology.

INTRODUCTION

Varieties of cellular processes are carried out by groups of proteins that interact with each other [1]. Understanding of these groups is a vital step to discover the complicated molecular interactions within cells [2, 3], and the abnormalities in protein interactions often lead to disease phenotypes. Recently, exploration of interactions between proteins encoded by disease genes in the human protein-protein interaction (PPI) network has been considered as one of the major and powerful approaches to reveal the molecular mechanisms underlying the complex diseases [4-6]. Therefore, the experimental detection of protein–protein interactions has been regarded as one of the major research field in system biology with promising applications in medicine [7, 8].

Various PPI networks have been constructed from compiling the data of studies which designed to characterize experimentally protein–protein interactions [9-12]. In order to get biological information, several networks' properties have been investigated. Network topology has a prominent role in understanding network architecture and performance [13]. Indeed, the organizational principles of biological systems have been described by

analyzing of topological properties of PPI networks [13, 14]. The results of several studies indicated that diseases proteins have some differences in topological properties compared with the non-diseases proteins [15-17].
Different experimental techniques have been recently used to detect physical interactions between proteins [18, 19]. It has been in turn obtained an enormous protein–protein interaction data from many small-scale designed studies, as well as from genome-scale investigations in organisms such as bacteria, yeast, worm, fly, and human [11, 20-23]. Therefore, the development of a number of specialized databases has been prompted to curate and maintain PPI data from the scientific literature and facilitate their availability for scientific community [24]. These databases have independent annotation efforts according to range of their research interests, which resulted in complementary as well as redundant information. It is worth noting that the different databases apply different rules for capturing the data and often use different systems for cross-referencing genes and proteins across biological databases [25, 26].
Considering that, the adoption of the Proteomics Standards Initiative- Molecular Interaction (PSI-MI) which controlled vocabulary and data structure has been regarded as a major step [27] to create a common framework for representing PPI data. Although all major PPI databases follow the PSI-MI standard principles, it is only the first stage. In parallel to this, the IMEx (the International Molecular Exchange) consortium committed to further consolidating the PPI data representations and curation policies [28].
These standardization efforts have noticeably facilitated the merging of PPI data, but this goal are impeded by the different strategies adopted by the databases [28]. For the human interctome, various PPI databases which report only experimentally verified interactions include protein interaction sets of limited overlap. These discrepancies could arise from some points such as different literature mining criteria, differences in PPI incorporation rates from small scale experiments, as well as differences in methods for PPI selection, curation and updating [25, 28-31]. Therefore, because of inherent heterogeneities in PPI databases, the use of PPI data to study topological features need more considerations. More worrying is about the way in which PPI information have been combined from multiple sources to reach more complete interactome, because topological features of studied proteins might undergo a considerable bias in this strategy.
The object of this study is to compare the topological properties of common human proteins between datasets derived from five experimentally deposited databases which downloaded from Cytoscape [32]; they are included of DIP [33], MINT [34], HPRD [35], IntAct [36], BIND [37]. To this end, we used UniProt datasets, downloaded from Cytoscape, as a validated reference set of nodes to select candidate ones. Two criteria were considered to select such nodes; first, the nodes had cumulative degree in UniProt PPI dataset and the second, they were common in all five PPI databases. We analyzed the topological features of those proteins in five datasets using statistical analysis such as calculation of Correlation coefficients to compare these features between our selected datasets. We found the different level of significant correlations (weak, medium and one robust) for selected human proteins' topological features between pair-wise relevant datasets.

METHODS

PPI databases and Proteins

In order to analyze the topological features of human PPIs, we first needed sources of protein-protein interactions data. We selected five databases which report only experimentally verified interactions. Via the Cytoscape version3.2, the open source network visualization and analysis software, four PPI databases were downloaded: DIP (The Database of Interacting Proteins), MINT (Molecular Interaction database), IntAct (the IntAct molecular interaction database), and BIND (The Biomolecular Interaction Network Database). Since HPRD (Human Protein Resource Database) database was not uploaded on Cytoscape, we downloaded it from the original database. At the next step, we selected thirty proteins which were common in all five databases. To obtain candidate proteins, UniProt dataset was used that involved verified protein interactions (downloaded via Cytoscape). Two points were considered for selection of the candidate proteins from UniProt. The first point was the proteins' degree as the proteins which selected from UniProt dataset had cumulative degree. It means that finally, the proteins have been ranked in increasing order according to their number of interactions. The second, the proteins were common in the five databases.

Selected topological features and Statistical analysis

Seven topological properties were selected to compare between PPI datasets: 1) the node degree (Ki) represents the number of links to node i. 2) Average shortest pathlength (ASP) indicates average shortest path between a node and all the nodes in the PPI network. The Betweenness centrality of node v is calculated as:

$$C_B(v) = \sum_{s \neq t \neq v \in V} \frac{\rho st(v)}{\rho st} \quad (3)$$

Where the number of all shortest paths between vertex s and t regarded as ρst, and the number of shortest paths which passing through a node v out of ρst is ρst(v). Indeed, this formula represents the ratio of the number of shortest paths passing through vertex v to the number of all shortest paths between s and t. The closeness centrality Cc (v) of a node v is calculated as:

$$C_c(v) = \frac{1}{\sum_{u \in v} dis(u,v)}) \quad (4)$$

The reciprocal of the total distance from a vertex v to all the other vertices in a graph is defined as Closeness centrality. The clustering coefficient C (v) of a node v is calculated as:

$$C(v) = \frac{2n_v}{k_v(k_v - 1)} \quad (5)$$

Where the number of links which connecting the kv neighbors of node v to each other defined as nv. To better understanding, in this formula, nv reveals the number of triangles which pass via node v, and kv (kv − 1)/2 indicates the total number of triangles which could pass via the node v. Eccentricity of a node v is calculated as:

$$C_E(v) = \frac{1}{\max\{dis(u,v): u \in V\}} \quad (6)$$

This formula implies that eccentricity E (v) of a vertex v is the greatest distance between v and any other vertex in a graph. The distance of a vertex from the center of a graph is represented by eccentricity of that vertex. The topological coefficient T (v) of a node v is calculated as:

$$T(v) = \frac{avg(S(u,v))}{g(u)} \quad (7)$$

where v is defined as a node that shares at least one neighbor with node u, and number of neighbors shared between node v and node u is represented by the function S (u,v), plus one if there is a direct link between node v and node u, and g(u) is the number of immediate neighbors node u. The topological coefficient of a node indicates to what extent this node shares neighbors with other nodes [38, 39].

For statistical analysis, correlations of seven topological features of proteins between five databases were evaluated by determining Spearman rank correlation coefficients. Both P-value and correlation coefficients were taken in to account to determine the correlation.

RESULTS

General features of databases and their analyzed average topological properties

In this study, we focused on five databases which considered as major literature-curated sources of PPIs and downloaded them from Cytoscape. As a first step, some general features of the databases were obtained from their respective websites (see Table 1). It should mention that BIND database is no longer active, and the relevant statistics acquired from BIND-translation project and iRefWeb. It seems that original PPIs of BIND database have been undergone an additional filter according to the goals of aforementioned websites. The number of human proteins and interactions which deposited on the individual databases was not the same with the ones downloaded from Cytoscape. It seems this difference is a result of Cytoscape filtering. Among mentioned databases, the most comprehensive database in terms of individual interactions is IntAct, with about 293,000 interactions from up to1082 different organisms. HPRD have been restricted to human proteins; it reports about 41,000 unique interactions, whereas IntAct reports about 102,000 unique human protein-protein interactions.

Table 1: General features of PPI databases.

Database	Number of proteins	Number of Interactions	Number of Organisms	Number of Experiments	Number of human proteins	Number of human Interactions	Number of human proteins in Cytoscape
DIP	26743	77514	665	74901	4110	6794	5115
MINT	35553	241458	362	16305	8762	26830	6679
HPRD	9673	41,327	1	20,164	9673	41,327	9618
IntAct	82947	292211	1082	32922	15676	101807	13754
*BIND	66754 (33177)	183495 (60770)	1588	unknown	8269 (5347)	15739 (9593)	11837

** The statistics are from BIND translate project and iRefweb*

Topological features of proteins in the databases were separately analyzed, and seven ones were selected for our five datasets (Supplementary file 1). We compared the average of the topological features in each dataset. The results have been represented on Table 2. The average of each topological feature is nearly in the same range at every dataset.

Table 2: Average of the topological features for databases

	DIP	MINT	HPRD	IntAct	BIND
Degree	5.1	4.6	8.2	7.1	3
Betweenness	0.008	0.003	0.003	0.002	0.014
Ave shortest pathlength	4.12	3.89	4.07	3.37	4.24
Closeness centrality	0.26	0.24	0.26	0.27	0.24
Clustering coefficient	0.09	0.1	0.11	0.09	0.05
Eccentricity	9.53	8.32	9.42	7.41	10.9
Topological coefficient	0.2	0.19	0.2	0.16	0.18

Correlation of topological features

To get more information about the organization of PPIs databases, the correlations of proteins' topological features between each dataset, regarded as a sample of relevant database, were calculated. Figure 1 shows the pair-wise correlations of seven topological properties between five selected datasets, and Scatter plots in figure 2 also were delineated to show the results for degree as illustrated forms (see also supplementary file 2). In case of degree, as the most elementary and interpretable topological feature, we find only one nearly robust significant correlation between HPRD and BIND (Spearman's rank r = 0.79, P = 0) and some medium significant correlations between DIP and IntAct (Spearman's rank r = 0.66, P = 0), MINT and HPRD (Spearman's rank r = 0.63, P = 0), MINT and BIND (Spearman's rank r = 0.53 P = 3 × 10-3) and IntAct and BIND (Spearman's rank r = 0.51 P = 5 × 10-3). There is also nearly weak significant correlation between HPRD and IntAct (Spearman's rank r = 0.39, P = 33 × 10-3). Betweenness is another topological feature for which also discovered three medium significant correlations between DIP and IntAct (Spearman's rank r = 0.56, P = 1 × 10-3), MINT and HPRD (Spearman's rank r = 0.67, P = 0) and HPRD and IntAct (Spearman's rank r = 0.58, P = 1 × 10-3). Besides, one nearly weak significant relation is between HPRD and BIND (Spearman's rank r = 0.47, P = 9 × 10-3). For the average shortest Path Length, we also get the following information: a medium significant relationship between DIP and IntAct (Spearman's rank r = 0.61, P = 0) and two nearly weak significant relations between MINT and IntAct (Spearman's rank r = 0.43, P = 19 × 10-3) and HPRD and IntAct (Spearman's rank r = 0.45, P = 13 × 10-3). The results for closeness centrality also show three nearly weak significant correlations between DIP and IntAct (Spearman's rank r = 0.47, P = 1 × 10-2), MINT and IntAct (Spearman's rank r = 0.41, P = 27 × 10-3) and HPRD and IntAct (Spearman's rank r = 0.46, P = 11 × 10-3). Clustering coefficient is the only

topological property which has one medium significant correlation between HPRD and IntAct (Spearman's rank r = 0.53, P = 3 × 10-3). Eccentricity as another property has only three nearly weak significant correlations between DIP and HPRD (Spearman's rank r = 0.43, P = 18 × 10-3), DIP and IntAct (Spearman's rank r = 0.42, P = 22 × 10-3) and MINT and IntAct (Spearman's rank r = 0.37, P = 45 × 10-3). Two medium significant correlations are found for topological coefficient between DIP and IntAct (Spearman's rank r = 0.56, P = 2 × 10-3) and HPRD and IntAct (Spearman's rank r = 0.56, P = 1 × 10-3), and two nearly weak significant correlations for DIP and HPRD (Spearman's rank r = 0.39, P = 37 × 10-3) and MINT and HPRD (Spearman's rank r = 0.42, P = 21× 10-3).

The findings indicate that degree has the maximum number of significant correlations between different datasets compared with other topological features, and clustering coefficient with only one significant correlation has the minimum number of significant correlations. The correlation of degree between HPRD and BIND is the only robust correlation. IntAct shows the most number of correlations for all studied topological features with other datasets, wherease BIND has the lowest number of correlations with others. Among IMEx contributors, IntAct and DIP have more number of significant correlations (four medium and two weak significant correlations), but MINT has only three weak significant correlations with IntAct. Althogh databses IntAct and HPRD have not followed the same curation policies, they show some significant correlations (three medium and three weak significant correaltions) for six topological features.

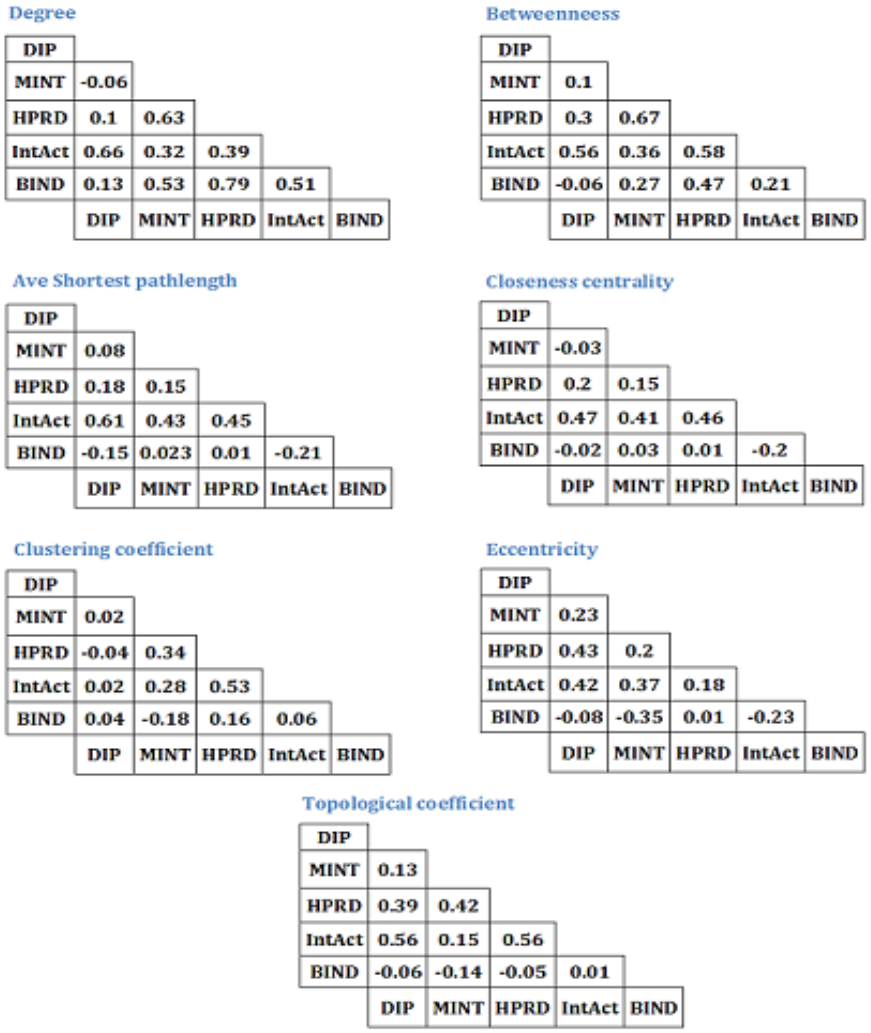

Figure 1: Pair-wise Spearman correlation coefficients (r) of seven topological features of common proteins in five datasets derived from their respective databases

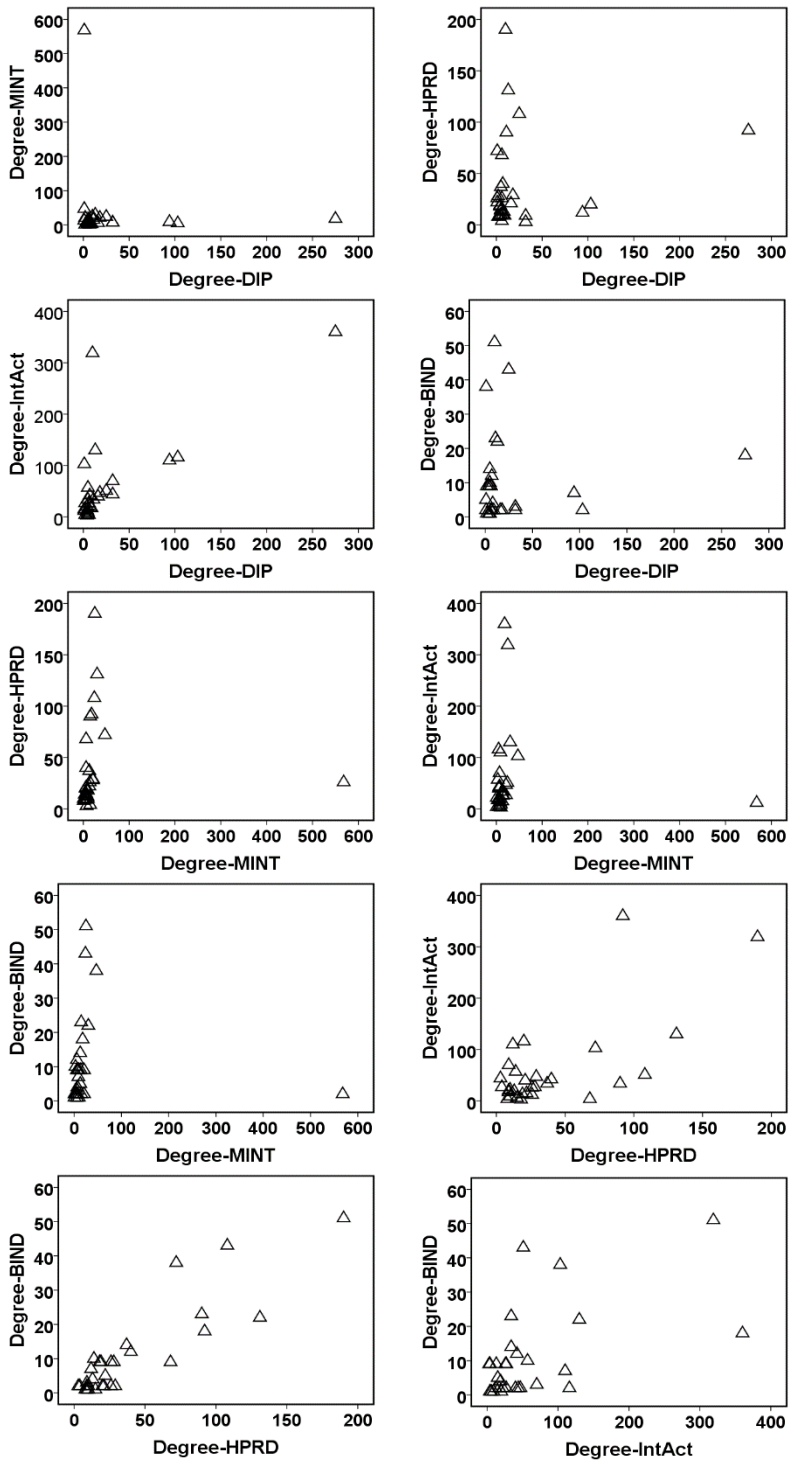

Figure 2: The comparison of degree Scatter plots between five datasets. Plots are representative of pair-wise correlations of shared proteins' degree between five databases: DIP, MNT, HPRD, IntAct and BIND

DISCUSSION

Protein-protein interaction networks have been used to gain insight into diseases mechanisms [40, 41], to determine drug targets [42] and discover novel network-based biomarkers [43]. The careful interpretation of PPI

data in various databases can provide biologically relevant conclusions [44]. In this study, we compared proteins' topological characteristics between derived datasets of five widely used public PPI databases. In doing so, we investigated the correlations of seven topological features in our five selected datasets; there were different significant levels of topological correlation (weak, medium and one robust) between datasets.

The five databases compared here as resources of our datasets, have different number of proteins and interactions. The mention of some points would shed light on such disparities about the size and entries of these PPI databases. To start with, most of these resources are independently funded and pursue their goals in isolation [28], and different curation policies have been implemented in each database. For instance, major PPI databases such as HPRD, BioGRID and IntAct have different research interests, which result in mainly independent annotation attempts; consequently, they contain complementary as well as redundant PPIs information [25]. Further to this, there are some noticeable issues overwhelming the representation of PPIs. The enormous space of estimated human interctome which comprises nearly 130,000 – 650,000 protein interactions [45, 46], and an inherently dynamic entity of interactome that varies among tissues, disparate cell process and environmental conditions [47] play considerable roles in this scenario. Thus, it seems important to define and enforce standards on the data formats to easily gather new PPIs information. Consistent with this, IMEx consortium agreed to distribute selected journals, as references of curation, among the members and determined standards for curation to increase curation coverage [28]. DIP, MINT and IntAct have created this consortium. HPR has not followed the consortium standards, but stays as largest human repositories, and BIND include informative piece of data that have been only deposited on this database; however, it is no longer active.

The noticeable point in our results was that, all studied topological features of common proteins, except clustering coefficient, have correlation in two databases, DIP and IntAct, which both follow the IMEx common curation policies. Although, MINT database is another major contributor of this consortium, there were only three significant correlations for topological features of common proteins between MINT and IntAct dataset and no significant correlation between MINT and DIP dataset.

Generally, we found nearly weak and medium significant correlations for seven topological features of shared proteins between five datasets. In this point of degree, significant correlations were more medium (four medium versus one nearly robust and one weak significant correlations) between all datasets. Among IMEx contributors, DIP and Intact had medium significant correlation, but MINT did not have significant correlations with DIP and IntAct; MINT had medium significant correlations with HPRD and BIND databases which both have followed different curation approaches in terms of MINT database. IntAct had a medium significant correlation with BIND and nearly weak correlation with HPRD. The only robust significant correlation (0.79) of degree was found between HPRD and BIND.

It seems different strategies have been implemented to organize PPIs in each database. In view of this, databases have different approaches to describe reported interactions from different species. For instance, HPRD is a human-centric database that modeled interactions onto human, whereas IMEx databases reported the exact protein species in the used experiment. In addition, the depth of curation also differs in databases as some databases may partially curate the literature to extract only the content of their special area of interest, like InnateDB and HPRD [26]. A recent report suggested that the agreement between two databases was about 42% for curated interactions and 62% in protein identifications when they used the same publication as a curation resource. Some difficulties have been mentioned for this matter : divergent organism assignments, the use of alternative protein identifiers and different representations of complexes [25]. With respect to the mentioned issues, our results were not far from the expectance in which some medium significant correlations of degree were found between different datasets.

The rest of topological features showed only a few numbers of medium and weak significant correlations in five datasets. DIP and IntAct had medium significant correlations in three features: betweenness, average shortest pathlength and topological coefficients, but they had two nearly weak correlations in closeness centrality and eccentricity. MINT, as another IMEx contributor, had three nearly weak significant correlations for average shortest pathlength, closeness centrality and eccentricity with only IntAct database. Besides, MINT had one medium significant correlation of betweenness and one nearly weak correlation of topological coefficient with HPRD. IntAct and HPRD showed three medium significant correlations in betweenness, clustering coefficient and topological coefficient, and they also had two nearly weak correlations for average shortest pathlength and closeness centrality. DIP and HPRD had two nearly weak correlations for eccentricity and topological coefficient. BIND had only one nearly weak correlation of betweenness with HPRD. Overall, two paired databases, DIP- IntAct and HPRD- IntAct had the more number of significant correlations for seven topological

features, whereas BIND had the least number of topological features correlations with others. It may be related to this matter that BIND database is no longer active. Lastly, the results emphasized this point that topological features in human PPI network are network-dependent.

Although the use of PPI data is so helpful to get invaluable information about molecular biology, the wrong use of these data may lead to fallacious conclusions. The analysis of topological features is so informative about the network structure, which help to elucidate many molecular processes underlying diseases. For instance, hub proteins are often targeted for the identification of possible lethal genes [48, 49], the development of novel drugs [50] or network disruption [51]. The study of betweenness as another feature suggests that even proteins with few interaction partners occupy important intermediate positions in network [52]. Therefore, the more reliable results would be convened from the study of Proteins' topological features if these features undergo fewer biases under the data assembling policies in the database. Since PPI databases are incomplete, some studies merged PPI databases to investigate the topological features [15, 17] and some did not [53]. To our knowledge, although the merge of PPIs on databases increase the number of interacting proteins, it may also increase the bias in topological features. Therefore, it should take into account for selection of PPI data in order to study topological characteristics.

CONCLUSIONS

Human PPI data can provide informative discretions of biological processes within cells, and due to this fact their applications have elevated in various aspects of biomedical researches. However, taking care of their accuracy is very important. Human PPI networks can be captured from some major experimentally deposited PPI databases which have not followed the same approaches to obtain PPI data. Therefore, this issue should be considered as an impediment to merge human PPI data form different databases in order to study network structure especially their topological features. Our results showed that there is more number of weak correlations of proteins' topological features between these public databases than medium ones; however, one significant strong correlation of degree has been seen between two PPI databases. It can be concluded that human proteins' topological features in current PPI databases are relatively dependent on the used databases. Although integration of different human PPI data might be an advantage because of increased coverage, it may result in more ambiguous interpretation of proteins topological features. Our results may help to select one database or more than one even if merging is your opinion.

ACKNOWLEDGMENTS:

This work was supported by the grant from the Proteomics Research Center in Faculty of Paramedical Sciences at Shahid Beheshti University of Medical Science.

Author Disclosure Statement:

The authors declare that they have no competing interests or financial disclosures.

REFERENCES

[1] Alberts B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell. 1998; 92: 291-294.

[2] Hartwell LH, Hopfield jj, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999; 402: C47-C52.

[3] Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS computational biology. 2007; 3: e59.

[4] Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein–protein interactions. Journal of medical genetics. 2006; 43: 691-698.

[5] Sun J, Jia P, Fanou AH, Webb BT, Van den Oord EJ, Chen X, et al. A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases–schizophrenia as a case. Bioinformatics. 2009; 25:2595-6602.

[6] Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterology and Hepatology from bed to bench. 2014; 7:17-31
[7] Lim J, Hao T, Shaw C, Patel AJ , Szabó G, Rual JF, et al. A protein–protein interac-tion network for human inherited ataxias and disorders of Purkinje cell degenera-tion. Cell. 2006; 125: 801-814.
[8] Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási Al. The human disease network. Proceedings of the National Academy of Sciences. 2007; 104: 8685-8690.
[9] Ito T, Chiba T, Ozawa, R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the Nation-al Academy of Sciences. 2001; 98: 4569-4574.
[10] Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, et al. A protein interac-tion map of Drosophila melanogaster. Science. 2003; 302: 1727-1736.
[11] Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot, A, Li N, et al. To-wards a proteome-scale map of the human protein–protein interaction network. Nature. 2005; 437:1173-1178.
[12] Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, et al. Proteo-me organization in a genome-reduced bacterium. Science. 2009; 326: 1235-1240.
[13] Zhu X, M Gerstein, M Snyder. Getting connected: analysis and principles of biologi-cal networks. Genes & development.2007; 21:1010-1024.
[14] Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature Reviews Genetics. 2004; 5:101-113.
[15] Sun J, Zhao Z. A comparative study of cancer proteins in the human protein-protein interaction network. BMC genomics. 2010; 11: S5.
[16] Goñi J, Esteban FJ, de Mendizábal NV, Sepulcre J, Ardanza-Trevijano S, Agirrezabal I, Villoslada P. A computational analysis of protein-protein interaction networks in neurodegenerative diseases. BMC systems biology. 2008; 2: 52.
[17] Choura M, Rebaï A. Topological features of cancer proteins in the human NR-RTK interaction network. Journal of Receptors and Signal Transduction. 2012; 32: 257-262.
[18] Brückner A, Polge C, Polge C, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. International journal of molecular sciences. 2009; 10:2763-2788.
[19] Braun P, Gingras AC. Gingras, History of protein–protein interactions: From egg-white to complex networks. Proteomics. 2012; 12:1478-1498.
[20] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, et al. Global land-scape of protein complexes in the yeast Saccharomyces cerevisiae. Nature.2006; 440: 637-643.
[21] Butland G, Peregrín-Alvarez J, Li J, Yang W, Yang X, Canadien V, et al. Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature. 2005; 433: 531-537.
[22] Guruharsha K, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N, et al. A Protein Complex Network of< i> Drosophila melanogaster</i>. Cell. 2011; 147:690-703.
[23] Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, et al. A census of human soluble protein complexes. Cell. 2012; 150:1068-1081.
[24] Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. Nucleic acids research. 2006; 34:D504-D506.
[25] Turinsky AL, Razick S, Turner B, Donaldson IM, Wodak SJ. Literature curation of protein interactions: measuring agreement across major public databases. Data-base. 2010; 2010: baq026.
[26] Orchard S. Molecular interaction databases. Proteomics. 2012; 12: 1656-1662.
[27] Kerrien S, Orchard S, Montecchi-Palazzi L, Aranda B, Quinn AF, Vinod N, et al. Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interac-tions. BMC biology. 2007; 5:44.
[28] Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, et al. Protein interac-tion data curation: the International Molecular Exchange (IMEx) consortium. Na-ture methods. 2012; 9:345-350.
[29] Klingström T, Plewczynski D. Protein–protein interaction and pathway databases, a graphical review. Briefings in bioinformatics. 2011; 12:702-713.
[30] Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, et al. Literature-curated protein interaction datasets. Nature methods. 2008; 6:39-46.
[31] Mathivanan S, Periaswamy B, Gandhi T, Kandasamy K, Suresh S, Mohmood R, et al. An evaluation of human protein-protein interaction data in the public do-main. BMC bioinformatics. 2006; 7: S19.

[32] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction net-works. Genome research.2003;13: 2498-2504.
[33] Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The database of interacting proteins: 2004 update. Nucleic acids research. 2004; 32: D449-D451.
[34] Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G .MINT: a Molecular INTeraction database. FEBS letters. 2002; 513:135-140.
[35] Prasad TK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database—2009 update. Nucleic acids research. 2009; 37: D767-D772.
[36] Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. Nucleic acids research. 2011; gkr1088.
[37] Bader GD, Betel D, Hogue CW. BIND: the biomolecular interaction network database. Nucleic acids research. 2003; 31: 248-250.
[38] Zhang A. Protein interaction networks: computational analysis. Cambridge Uni-versity Press; 2009
[39] Wang Z, Cao R, Taylor K, Briley A, Caldwell C, Cheng J The properties of genome conformation and spatial gene interaction and regulation networks of normal and malignant human cell types. PloS one. 2013; 8: e58793.
[40] Feldman I, Rzhetsky A, Vitkup D. Network properties of genes harboring inherited disease mutations. Proceedings of the National Academy of Sciences. 2008; 105: 4323-4328.
[41] Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PloS one. 2011; 6: e20284.
[42] Zhu M, Gao L, Li X, Liu Z, Xu C, Yan Y, et al. The analysis of the drug-targets based on the topological properties in the human protein-protein interaction net-work. Journal of drug targeting. 2009; 17: 524-532.
[43] Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature biotechnology. 2009; 27: 199-204.
[44] Chatr-aryamontri A, Ceol A, Licata L, Cesareni G. Protein interactions: integration leads to belief. Trends in biochemical sciences. 2008; 33: 241-242.
[45] Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf Ca. Estimat-ing the size of the human interactome. Proceedings of the National Academy of Sciences. 2008; 105: 6959-6964.
[46] Venkatesan K, Rual JF, Vazquez A , Stelzl U, Lemmens I, Hirozane-Kishikawa T, et al. An empirical framework for binary interactome mapping. Nature methods. 2008; 6:83-90.
[47] Ideker T, Krogan NJ. Differential network biology. Molecular systems biology. 2012; 8(1).
[48] Jeong H, Mason SP, Barabási A-L, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001; 411: 41-42.
[49] Coulomb S, Bauer M, Bernard D, Marsolier-Kergoat MC. Gene essentiality and the topology of protein interaction networks. Proceedings of the Royal Society B: Bio-logical Sciences. 2005; 272:1721-1725.
[50] Hase T, Tanaka H, Suzuki Y, Nakagawa S, Kitano H. Structure of protein interac-tion networks and their implications on drug design. PLoS computational biology. 2009; 5: e1000550.
[51] Quayle AP, Siddiqui AS, Jones SJ. Perturbation of interaction networks for applica-tion to cancer therapy. Cancer informatics. 2007; 5: 45.
[52] Joy MP, Brock A, Ingber DE, Huang S. High-betweenness proteins in the yeast protein interaction network. BioMed Research International. 2005; 2005: 96-103.
[53] Wu C, Zhu J, Zhang X. Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes. BMC bioinformatics. 2012; 13: 182.