

Research Article

Drug Target Protein-Protein Interaction Networks: A Systematic Perspective

Yanghe Feng,¹ Qi Wang,¹ and Tengjiao Wang²

¹Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China

²Department of Bioinformatics, Second Military Medical University, Shanghai, China

Correspondence should be addressed to Tengjiao Wang; tjwang@smmu.edu.cn

Received 15 January 2017; Revised 9 March 2017; Accepted 10 May 2017; Published 11 June 2017

Academic Editor: Yudong Cai

Copyright © 2017 Yanghe Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The identification and validation of drug targets are crucial in biomedical research and many studies have been conducted on analyzing drug target features for getting a better understanding on principles of their mechanisms. But most of them are based on either strong biological hypotheses or the chemical and physical properties of those targets separately. In this paper, we investigated three main ways to understand the functional biomolecules based on the topological features of drug targets. There are no significant differences between targets and common proteins in the protein-protein interactions network, indicating the drug targets are neither hub proteins which are dominant nor the bridge proteins. According to some special topological structures of the drug targets, there are significant differences between known targets and other proteins. Furthermore, the drug targets mainly belong to three typical communities based on their modularity. These topological features are helpful to understand how the drug targets work in the PPI network. Particularly, it is an alternative way to predict potential targets or extract nontargets to test a new drug target efficiently and economically. By this way, a drug target's homologue set containing 102 potential target proteins is predicted in the paper.

1. Introduction

Drug target studies have been conducted in both dry and wet labs from experimental designs to target identification and validation steps. There are two ways for target discovery: a system approach and a molecular approach [1, 2]. In 2006, Imming et al. [3] catalogued 218 molecular targets for approved drug substances and Overington et al. [4] proposed a consensus number of 324 drug targets for all classes of approved therapeutic drugs. Rask-Andersen et al. [5] analyzed trends in the introduction of drugs that modulate previously unexploited targets and discussed the network pharmacology of the drugs in our dataset. Yildirim et al. [6] built a bipartite graph composed of US Food and Drug Administration-approved drugs and proteins linked by drug target binary associations. By using chemical 2D structural similarity, Keiser et al. predicted new molecular targets for known drugs [7]. Campillos et al. [8] used phenotypic side-effect similarities to infer whether two drugs share a target.

The main molecular targets for drugs are proteins (mainly enzymes, receptors, and transport proteins) and nucleic acid (DNA and RNA). It is indeed gratifying that development of research in cell biology, molecular biology, and biochemistry produced a remarkable compendium of knowledge on the function and molecular properties of individual proteins [9]. However, a protein usually carries out a typical function by regulating other molecules. In other words, it rarely acts alone. Therefore, in the past few years, with the emergence of high-throughput technologies for omics data, like yeast 2 hybrid protein interactions, research of protein-protein interactions (PPI) has been occurring with high frequency. Bultinck et al. [10] concluded that a growing number of functional PPI modulators are being reported and clinically evaluated. PPI provides us with more information for understanding the relationship between drug targets and other proteins in a systematic point of view.

Many researchers believe that topographical analysis of the complex network of intercellular protein interactions may

also lead to new avenues for target prediction [11]. Based on the graph modeling theory in computer science, protein networks are extensively being studied in the system biology field. Royer et al. [12] used power graph analysis to explicitly represent reoccurring network motifs. In a systematic point of view, the gene regulatory network and drug targeted protein network are different from their normal counterparts. Currently, the interactions between drugs and targets have also been studied as an interactive network. Chautard et al. [13] described the high-throughput methods used to identify new interactions and to build large datasets that record the identified interactions. Yamanishi et al. [14] characterized four classes of drug–target interaction networks and revealed significant correlations between drug structure similarity, target sequence similarity, and the drug–target interaction network’s topology. The essential intentions of these approaches are trying to take the time-specific or space-specific information into account. Thus, these systems biology approaches will lead to time-sensitive, space-sensitive, and synergistic treatments taking the multidimensional use of drugs into consideration [13].

These studies mainly focused on interactions of drugs targets rather than the targets’ topological features. It is a reasonable way to identify targets on which the typical drugs work. However, the targets topological features are helpful to predict new targets because most of them have similarity on some topological features which are different from normal proteins. To construct a panoramic view of drug targets, in this paper, we examined the three main institutive views about drug target characteristics: intermediaries, source of the drug stimulus, and special topological features. Based on the PPI network, we analyzed lists of the topological indices related to the three traditional views above. The results show, somewhat surprisingly, that the topology of a drug target is not to help it as intermediary or be the source of the drug stimulus. On the other hand, drug target proteins indeed have some special topological features that are significantly different than normal proteins.

The remainder of this paper is organized as follows. The Data Collection elaborates the details about data sources of drug target properties and the PPI network used in the paper. The analysis on drug target protein topology under three transitional points is in the Analysis. The Discussion discusses the drug target’s topological characteristic and its applications. This paper concludes in the Conclusion with the impact of the paper and our future work insights.

2. Data Collection

Proteins are the main catalysts, structural elements, signaling messengers, and molecular machines of biological tissues [15]. The interactions between proteins form the basis for signal transduction pathways and transcriptional regulatory networks. Therefore, it is very important in harmonizing the events in a cell. Target proteins are functional biomolecules that are addressed and controlled by biologically active compounds. Currently, the main resource of protein interactions is from five most widely used PPI databases (HPRD version R9) [16], IntAct (2010-02-17 downloaded) [17], BioGRID

(version 2.0.63) [18], MINT (version 2010-05-05) [19], and DIP (version 2009-12-30) [20]. The databases contained a combined 65,785 nonredundant interactions.

In this paper, the main information of drug targets was extracted from the DrugBank database, in which the approved targets set (version 3.0) [21] contains 1,604 proteins. Then PISCES [22] was used to remove those sequences with an identity larger than 20% for both the drug target and the nontarget sequence. Using this method, we gained 517 drug targets and 3,834 common proteins. The drug protein properties included single peptide cleavage [23], transmembrane helices [24], low complexity region [25], N-glycosylation [26], and O-glycosylation [27]. These chemical properties of the amino acids of proteins determine the biological activity of the protein. Therefore, they can provide us with information that whether a protein is suitable to be a drug target protein. The chemical and physical properties which were used in our analysis are 26 amino acids (counted by Mole%) and number of charged residues, basic residues, acidic average molecular weight, and isoelectric point are used to train the model and predict the potential drug targets [28]. Besides, they are important clues as well as the PPI topological features for the judgement of which proteins could be targets. Here we use pepstats, an online software from EMBOSS [29] to calculate statistics of protein properties. A more detailed information on these properties including hypothesis test is in [30]. Finally, the proteins dataset contains 39 chemical and physical properties.

After integrating the DrugBank target protein data and the PPI data, we gained 1,361 proteins that have both network properties and chemical-physical properties. Of the 1,361 proteins, 149 are known drug targets and the remaining 1,212 are yet to be tested. There are 10,197 proteins without chemical and physical properties. Unfortunately, although there are improved methods to reduce the protein-protein interaction’s false positive rate [31, 32], the information of protein-protein interactions is always incomplete so that there are some isolated nodes (e.g., PDXP, ODZ1, NT5M, and IL17F) and small components that only contained two or three nodes (e.g., the component consisting of KLRG1 and LEPROTL1 and the component consisting of PCYT2 and JMJD5). Hereby, in order to reduce the effect from lack of complete interactions, we used the maximal connected component instead of the original PPI. It contains two types of proteins, 138 drug targets (D) and 11,163 pending test proteins (PT). The latter consist of 1,180 proteins with 39 chemical-physical properties (PT1) and 9,983 without any properties (PT2).

Table 1 shows the overall description of the collected data. It gives a summary of the data which were performed in our manuscript. It shows the fact that the number of known drug target proteins is far less than the number of pending test proteins. It reveals the feasibility of this analysis and the possibility that there are many potential drug targets in the pending test data. The known DT proteins which we collected from the DrugBank database were removed from non-DT proteins dataset. The proteins redundancy method [33] was performed to obtain our DT and non-DT protein dataset. The coverage ratio is calculated in the way using the used data

TABLE 1: Summary of data.

	Drug targets (D)	Pending test (PT)		Proteins (P)	Edges (E)	Drug targets ratio (DR = D/P)	Pending test ratio (PR = PT/P)
		PT1	PT2				
Original data	149	1212	10197	11558	65772	0.23%	99.77%
Used data	138	1180	9983	11301	65547	0.21%	99.79
Coverage ratio	92.6%	97.4%	97.9%	97.8%	99.7%	91.3%	100%

(the number of drug target or pending test proteins which is used in the analysis) divided by the original data (the number of drug target or pending test proteins acquired from the public dataset).

3. Drug Target Network Topology Analysis

As stated in classical complex networks theory, the drug targets network is represented as an undirected network $G = (V, E)$, where V denotes the protein in D set or PT set and E is the interactions between each proteins pair. In the paper, the drug targets network contains 11,301 nodes and 65,547 edges. For each node $i \in V, k_i$ denotes the degree of it. A is the adjacency matrix for the network, where $A_{i,j} = 0$ when there are no interactions (no edge) between nodes i and j . Similarly, $A_{i,j} = 1$ when there is an interaction between each other.

The drug target is the native protein in the body whose activity is modified by a drug resulting in a desirable therapeutic effect. Different drugs act on molecular targets at different locations in the cell. In the human body, all cells have membrane that enclose the cytoplasm. The cell membrane consists of two identifiable layers, each of which is made up of an ordered row of phosphoglyceride molecules such as phosphatidylcholine. A phosphoglyceride molecule consists of a small polar head group and two long hydrophobic chains. In the cytoplasm, there are several structures, one of which is the nucleus that acts as the control center for the cell. There are also many other structures within a cell, such as mitochondria, Golgi apparatus, and the endoplasmic reticulum. As the drug targets are the special proteins through which the drugs carry out their specific functions, they are thought institutively as (1) the intermediaries which play an important role on interactions of the drug targets network; (2) the sources which receive the drug stimulus and convert it into another stimulus that can be responded to by normal proteins; (3) the proteins which have special topological and functional significance. According to these three points, we analyzed the listed topological features of the drug targets network, including degree and betweenness for the intermediary function, eccentricity, and average distance for the source function, modularity, coreness, cluster coefficient, and eigenvector centrality for special topology. However, from analyses of the PPI topological indices, the drug targets do not have the first two characteristics. Actually, the results show they are similar with other proteins on intermediary and source functions. In comparison, there are some significant differences on special topology.

3.1. Drug Target as Intermediary. Diseases are regulated by complex biological networks and depend on multiple steps of genetic and environmental challenges to progress [34]. Disease-relevant intracellular PPI occurring at defined cellular sites possess great potential as drug targets. They permit highly specific pharmacological interference with defined cellular functions [35]. In other words, the drug targets seem to be the proteins through which the drug effects tend to be spread over the PPI to stimulate other related proteins. In this paper, we studied the degree and betweenness to analyze the intermediary function of the drug targets.

3.1.1. Degree. The degree of a node in a network (sometimes referred to incorrectly as the connectivity) is the number of connections or edges the node has to other nodes. In protein interaction networks, the hubs are defined as ones that have a higher degree than others; for example, Vallabhajosyula et al. [36] suggested that some of the highest degree proteins should be defined as hubs which have special topological and functional significance. We analyzed the degree distribution of the known drug targets and other proteins from Pending Proteins (PT) shown in Figure 1, where the x -axis means the degree and the y -axis means the proportion of the protein with that degree. The right subfigure, the double logarithm coordinate system, shows an obvious power-law distribution with significant long tail characteristics for both groups. The highest degree of the drug targets is 103 compared to 667 in PT. The average of degree in drug targets is 13.3 compared to 11.6 of the proteins in PT.

Figure 1 shows the degree distributions of two types of proteins during $k_i \in [1, 103]$ in which all of the known drug targets distribute. The inset is the overall view of both distributions. From Figure 1, the proteins with the highest degree in the drug targets network are usually not drug targets. On the contrary, the degree distributions are similar between drug targets and the majority of other proteins. They are following the well-known power law just with different parameters, which seems to be counterintuitive. The result implies that the drug target proteins are actually not "hub" proteins but rather they probably carry specific functions at a certain level of interactome as a common protein. Due to robustness and resilience of the PPI network based power law, the topological properties will not change if some small degree nodes are removed. But for the hubs, Vallabhajosyula et al. [36] referred that removal of the hubs would cause significant change for the PPI network. According to it, we analyze the change of degree after the drug targets are removed. There is only decreasing from 11.6 to 11.4 after those drug targets were removed. Therefore, the results imply that

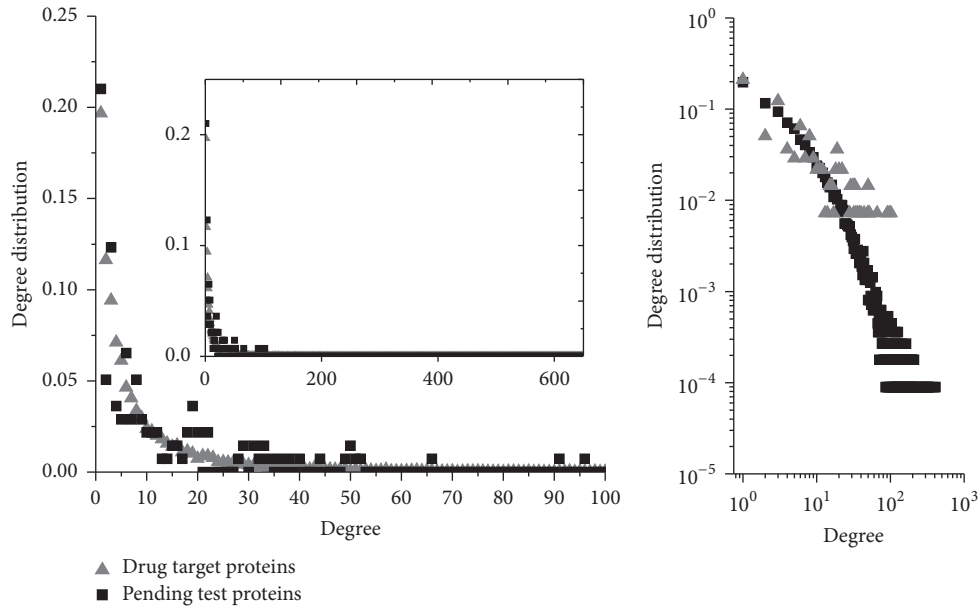


FIGURE 1: Degree distributions of two types of proteins. The inset is the overall view of both distributions. Drug targets and common proteins are following the well-known power law just with different parameters. The highest degree of the drug targets is 103 compared to 667 in PT. It indicates that the drug targets are not the hubs.

the degree distribution of the drug targets is not a significant topological network feature. Actually, it is similar to cancer research findings such as the findings of Barillot et al. [37] which propose a notion about a router that is not necessary to have a high connection like a hub. However, it always plays an important role of propagating the biological signal to a local hub or a global hub protein.

3.1.2. Betweenness. Betweenness is the number of times a node is in the shortest paths between two other nodes. In the PPI network, if the drug targets are the proteins that play an important role on intermediary, they may have higher betweenness than others. Actually, many studies show that some interactions are more important in nonhub proteins based on betweenness. For example, in the yeast proteins network, the coordinated functionality is carried out by the connectors which have high betweenness, even though they have low degree [38]. Particularly, in the PPI network which has undirected edges, Yu et al. [39] found the betweenness is more essential than degree with gene essentiality and expression dynamics. Therefore, we try to study the intermediary function of the drug targets based on betweenness.

The betweenness centrality of a node v is given by

$$\text{Btwn}(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (1)$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v .

The normalized betweenness (NB) can be calculated without a loss of precision

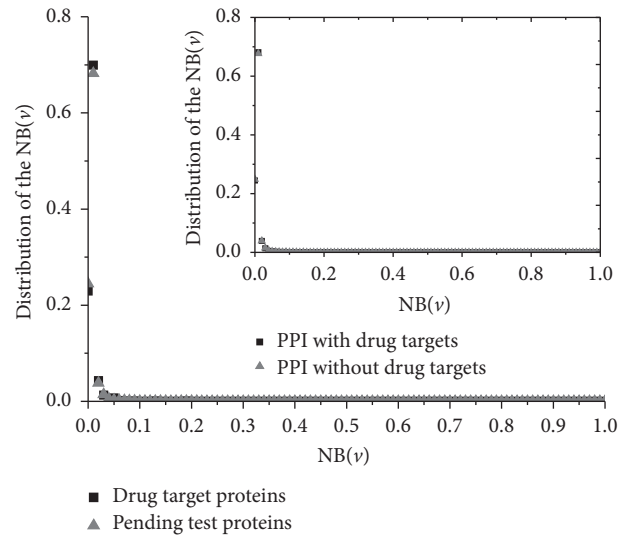


FIGURE 2: Distribution of normalized betweenness of drug target proteins and pending test proteins. The betweenness of drug target proteins is lower than others. The distributions of the two types of proteins are similar. The inset shows there are few changes between the betweenness distributions of the original PPI and the new PPI with the known drug targets removed.

$$\text{NB}(v) = \frac{\text{Btwn}(v) - \min(\text{Btwn})}{\max(\text{Btwn}) - \min(\text{Btwn})}. \quad (2)$$

It results in $\max(\text{NB}) = 1$ and $\min(\text{NB}) = 0$. Figure 2 showed the distributions of the drug targets and other proteins in the PPI network.

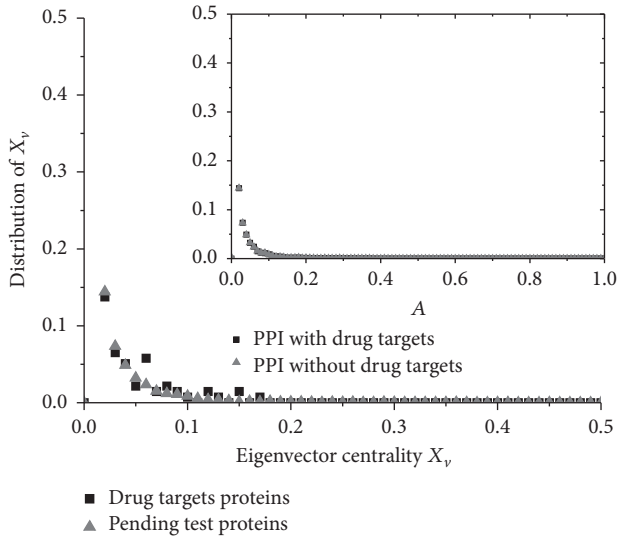


FIGURE 3: Distribution of eigenvector centrality. Drug targets and other proteins have the same distribution of the eigenvector centrality. The inset shows that the amount of changes on the eigenvector centrality after removing the drug targets is negligible. But the average eigenvector centrality of drug targets is not lower than others although the results of the analysis on degree and betweenness implied the drug targets are not the hubs.

In Figure 2, there is a surprising result as the conclusion from analysis of the degree. The drug target proteins do not have high betweenness but rather lower than most common proteins. The highest NB of the drug target proteins is BCL2_HUMAN which is only 0.0497. Even though the drug targets are all removed from the PPI network, the betweenness of the remaining proteins is nearly unchanged shown in inset of Figure 2. It also clearly shows that the drug target proteins hold the similar distribution of betweenness as other proteins. This implies that relevance of a drug protein as an organizing regulatory molecule is fairly the same as other proteins.

From the analysis on degree and betweenness, it implies that the drug targets are not the important proteins from the view of the connectivity of the PPI network. In a complex network, the eigenvector centrality is a measure of the influence of a node. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes [40, 41]. We examined whether the eigenvector centrality x_v of the drug target is lower and the distribution is similar to the normal proteins (Figure 3). The centrality score x_v of vertex v is defined as

$$X_{v_i} = \frac{1}{\lambda} \sum_{v_j \in M(v_i)} X_{v_j} = \frac{1}{\lambda} \sum_{v_j \in G} A(v_i, v_j) X_{v_j}, \quad (3)$$

where $M(v_i)$ is a set of the neighbors of v_i and λ is the greatest eigenvalue of the adjacent matrix A .

Eigenvector centrality measures the centrality of a node by judging how many nodes it connected to with high degree.

The distributions of the drug target proteins and other proteins are quite similar as shown in Figure 3. Even if the drug targets are removed, the change of the eigenvector centrality is so small that it can be ignored (see inset of Figure 3). It seems to be the same as what degree and betweenness imply. However, the average eigenvector centrality of drug targets is 0.022, which is not lower than normal proteins 0.02. This conclusion seems to oppose the conclusion that drug targets are not the important proteins from the view of the connectivity based on the degree and betweenness. It indicates that the drug targets have interactions with the high degree proteins although they are not hubs of the network as what degree and betweenness implied. Actually, it is the important feature of the drug targets and is analyzed again when we consider coreness of the PPI network.

Intuitively, the drug effects are spread by some special proteins called drug targets which have either high degree or high betweenness. Actually, in many studies on protein interactions or gene interactions, the hubs or the nodes with high betweenness are the important ones [42–44]. However, the conclusions of our analysis are opposite: the drug targets have lower degree and betweenness, but also there are no significant differences on the distributions of both topological indices between drug targets and other proteins. It implies that the intermediary is not the main function of the drug targets. There are many potential reasons for this. One reason is that the drug design method is based on the ligand and structure at present. Therefore, the drug’s stimulus is effective for some specific proteins rather than a large amount of proteins.

3.2. Drug Targets as Source. The drug actions depend on the complex signaling transduction networks of cells or the complicated profile of drug potency and selectivity [45]. In most cases, if the different targets stimulated by a drug, the in vivo effect on the signaling pathway should be changed and the drug’s efficiency to inhibit the activity (usually measured as phosphorylation level) will be different [46]. In other words, it is possible that a drug target seems to be the source which receives the drug stimulus and converts it into another stimulus that can be responded to by normal proteins. If so, the drug targets should be the sources from which it is convenient to access other proteins in PPI networks. Even if a protein has low degree and betweenness, it is possible for it to have short distances with others. More generally speaking, the convenience of the drug targets depends on the degree and betweenness analyzed above but also the distances to other proteins. In this paper, we studied the average distance and eccentricity to analyze source functions of drug targets.

3.2.1. Average Distance. Average distance, also called average path length, is a concept in network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. Here, we consider an unweighted graph G ; let $d(v_i, v_j)$ denote the shortest distance between v_i and v_j . Assume that $d(v_i, v_j) = 0$ if $i = j$ or D_G if v_i cannot be reached from v_j , where D_G is the diameter of the PPI network. Note that it is different with the traditional average distance which defines $d(v_i, v_j) = 0$ if v_i cannot be

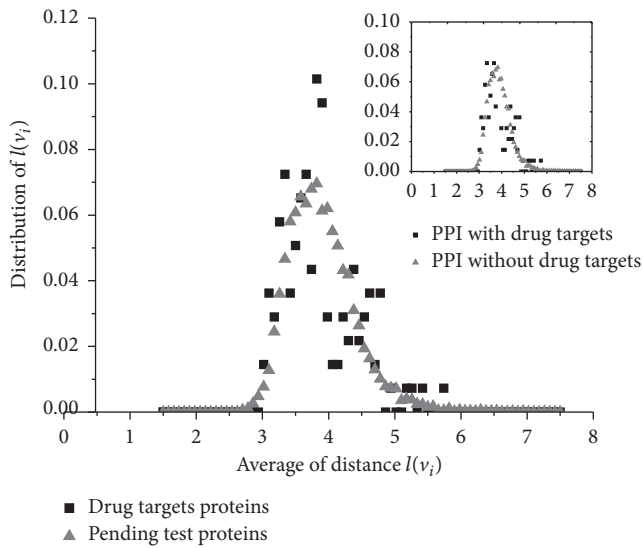


FIGURE 4: Distribution of average distance. The known drug target's distribution of average distance is similar with common proteins. The inset shows there are few changes on the average distance after the protein targets were removed from the PPI network.

reached from v_j . However, we are concerned about how a protein spreads the effects of the drug to others. If there are no interactions between two proteins, the stimulus on each drug is not transferred to the other. The distribution of the average distance $l(v_i)$ is shown in Figure 4.

The average distance shows the overall convenience of the proteins to communicate and/or affect their reciprocal function. It is also a sign of functional convergence [47]. In PPI networks, the special functions with low average distance cause many diseases. For example, when the cancer related proteins are compared with normal proteins, the average distances are lower [42]. But in our study, the drug targets do not have this feature. Figure 4 shows that the distribution of average distance is similar with other proteins. Meanwhile, even if the protein targets are removed from the PPI network, the distribution of average distance hardly changed. Hence, the drug targets stimulated by a special drug do not have more significant convenience of propagating the effects to other proteins.

3.2.2. Eccentricity. The eccentricity $\epsilon(v)$ of a vertex v is the greatest geodesic distance between it and any other vertex [48]. It can be thought of as how far a protein is from the proteins farthest from it in the graph [49]. Indeed, if the eccentricity of the node v is low, the other nodes are in proximity. On the contrary, if it is high, it implies that there is at least one node (and all its neighbors) that is far from node v . Let $d(v_i, v_j)$ denote the distance (number of edges connected) between vertices i and j ; then the eccentricity $\epsilon(v) = \max_{v_i, v_j \in V} d(v_i, v_j)$. Figure 5 shows the distribution of the eccentricity.

The eccentricity shows the easiness of a protein to be functionally reached by all other proteins in the network. Thus, a

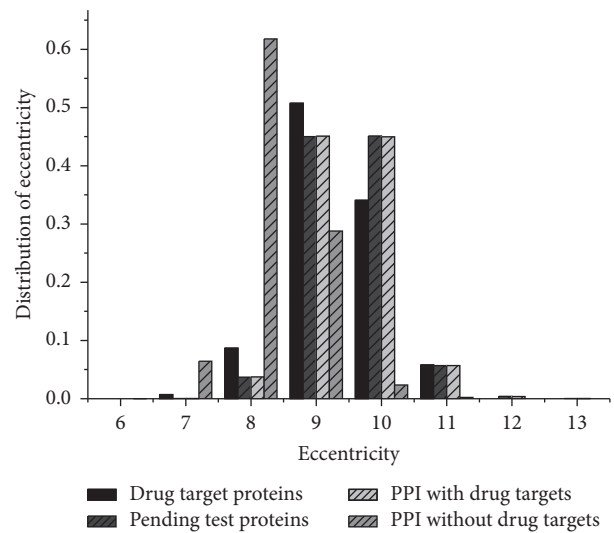


FIGURE 5: The distribution of eccentricity. It shows that the difference in eccentricity between drug targets and others is not remarkable although most targets' eccentricity is 9 and the normal proteins' eccentricity is 9 and 10. However, the eccentricity is changed apparently after removing the drug targets from PPI networks.

protein with low eccentricity is subject to a more stringent or complex regulation so that it could easily influence several other proteins. Similar to the average distance, there are few observed differences on the distribution of eccentricity between known drug targets and other proteins (Figure 5). However, after removing the drug target proteins from the original proteins network, the eccentricity of most proteins reduced to 8 compared with 9 and 10 in the network with drug targets (Figure 5). In a network, the eccentricity $\epsilon(v)$ should be increased if the hubs or some important connectors are removed. Hence, the results above including analysis of the intermediary showed, somewhat surprisingly, that the drug targets are not the proteins that play the important role on connectivity of the PPI networks. On the contrary, most of them are stimulated by drugs and this modification on their activities is spread to other proteins. Therefore we hypothesized that the drug mainly fulfills effects on some special targets, rather than stimulating other proteins through drug targets. For example, if avoiding the host's defense mechanisms and inhibiting nonspecific distributions in the liver and spleen by targeted drug delivery, a cardiac tissue system can reach the intended site of action in higher concentrations [50].

3.3. Drug Targets Topological Structure Characteristics. According to analysis on intermediary and source functions of the drug targets, there are few significant characteristics with respect to targets discovery. These results indicate that the function of spreading drug stimulus seems not to be as important as the traditional view for the drug targets. Actually, several studies show that drug effects only depend on some special proteins rather than impacting most of proteins. For a specific drug, the physical and chemical

properties of the target proteins are exactly important for response on the drug stimulus. Moreover, these proteins with different physical and chemical properties often have different topological features.

The process of finding a new drug against a chosen target for a particular disease usually involves high-throughput screening (HTS), wherein large libraries of chemicals are tested for their ability to modify the target [51]. For example, if the target is a novel GPCR, compounds will be screened for their ability to inhibit or stimulate that receptor (see antagonist and agonist): if the target is a protein kinase, the chemicals properties will be tested for their ability to inhibit that kinase. The physicochemical properties associated with drug absorption include ionization (pKa) and solubility; permeability can be determined by PAMPA and Caco-2. PAMPA is attractive as an early screen due to the low consumption of drug and the low cost compared to tests such as Caco-2, gastrointestinal tract (GIT), and Blood-brain barrier (BBB) with which there is a high correlation. Another important method for drug discovery is drug design, whereby the biological and physical properties of the target are studied, and a prediction is made of the sorts of chemicals that might, for example, fit into an active site. One example is fragment-based lead discovery (FBLD) [52]. Novel pharmacophores can emerge very rapidly from these exercises. In general, computer-aided drug design is often but not always used to try to improve the potency and properties of new drug leads.

Meanwhile, eigenvector centrality also implied that the drug targets may have the special structure that causes their roles on connectivity to be equal to or even more important than other proteins with higher degree and betweenness. Some structures are found to link with significant proportion of proteins. In the yeast proteins interaction networks, the protein pairs inside some special subnetworks corresponding to protein complexes tend to show interactions maps of specific biological processes [53]. In breast cancer prognosis, the changes of network modularity may be a defining feature of tumor phenotype that determines patient prognosis [54]. Moreover, the average clustering coefficient values of the cancer proteins interaction networks were lower. It implies that the proteins have a lower tendency to form clusters [42]. Similarly, the clustering coefficient of the drug targets (0.06) is lower than other proteins (0.12) in drug proteins interaction networks as well. Furthermore, most drug targets distribute during 0 to 0.1. As eigenvector centrality, it implies the drug targets may be in some special subnetworks. In this section, beside the physical and chemical properties, we examined which topological characteristics of the PPI networks contribute for response to the drug stimulus based on modularity and coreness.

3.3.1. Modularity. Modularity is the degree to which the components of the networks may be separated and recombined. The definition of modularity varies in different fields with similar essence. In biology, modularity refers to the concept that organisms or metabolic pathways are composed of modules. In the paper, we studied modularity Q of networks, which is a benefit function defined as (4) that measures the

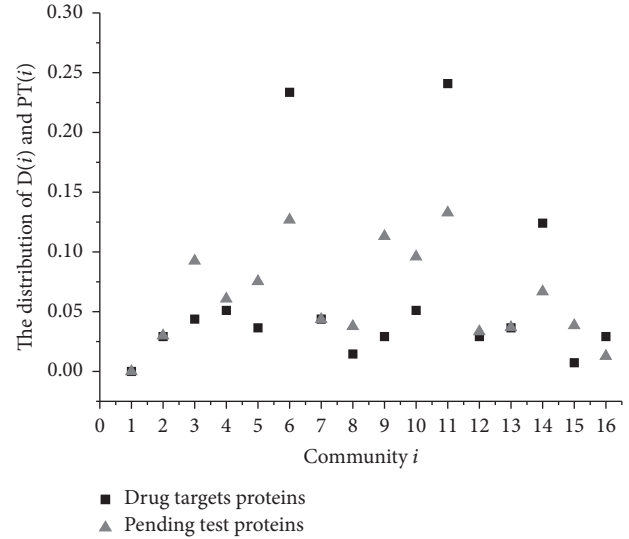


FIGURE 6: The distribution $D(i)$ and $PT(i)$ of the communities. There are 15 main protein communities consisting of 11,099 proteins detected in PPI. The majority of the drug targets are in the three communities 5, 10, and 13.

quality of a division of a network into groups or communities [55].

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{i,j} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j), \quad (4)$$

where the degree of node i assigned to community c_i is k_i , if there is no interaction between proteins i and j , and 1 otherwise, $m = (1/2) \sum_{i,j} A_{i,j}$, and $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise.

There are many effective algorithms to detect communities by maximizing the modularity based on (4). In the paper, we used the method in paper [56] and detected 34 communities. There are 15 main protein communities consisting of 11,099 proteins detected in PPI. The distributions $D(i)$ and $PT(i)$ of the communities are shown in Figure 6, where $D(i)$ is the percentage of the drug targets in community i to the amount of known drug targets. Similarly, $PT(i)$ is the percentage of the pending test proteins detected into community i .

However, if a community has most of the proteins, it also probably contains most of the drug targets. Such community should not be considered as a target-like community which mainly consists of drug targets. Furthermore, we calculate $rd(i) = D(i)/PT(i)$ as the degree to which community i is likely to be a target-like community. Because $D(i) = i_D/|D|$ and $PT(i) = i_{PT}/|PT|$, $rd(i) = (D(i)/PT(i))/(|D|/|PT|)$, where i_D and i_{PT} are the amounts of the drug targets and other proteins in community i , respectively, and $|D|$ and $|PT|$ are the amounts of the set D and of the set PT .

Therefore, $rd(i) > 1$ indicates that $i_D/i_{PT} > |D|/|PT|$. In other words, community i tends to be as a target-like community. Similarly, if $rd(i) < 1$, community i is closer to nondrug target proteins.

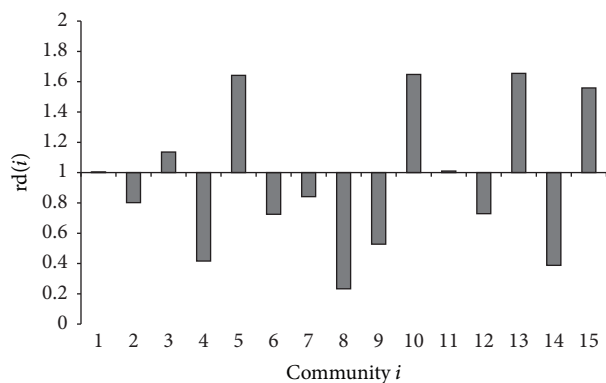


FIGURE 7: $rd(i)$ of the communities. When $rd(i) > 1$, it indicates that the community i tends to be a target-like community. Similarly, if $rd(i) < 1$, the community i is closer to nondrug target proteins.



FIGURE 8: The full view of communities 5 and 10.

Figure 6 shows 15 communities which contain the most drug targets. In particular, the known drug target proteins mainly gathered around communities 5, 10, and 13. These three communities account for almost 66% known drug targets. In contrast, the pending test protein is flatter and more spread out for several communities. It implies there are less nontarget data around the communities in which drug target data mainly clustered together. In particular, Figure 8 shows communities 5 and 10 in which the drug target proteins most mainly distribute. In Figure 8, the drug target is not always the one with highest degrees in either whole protein network or communities they mainly cluster.

These three communities also have highest $rd(i)$ in Figure 7. In addition, community 15 is the other significant target-like community as well. Meanwhile, the nondrug targets are clustered in communities 4, 8, 9, and 14. The results based on modularity imply there are target-like communities existing in the PPI networks. It is helpful to study which proteins are the potential drug targets and to understand how the drug targets work. Actually, some approaches have been explored for drug discovery by specific structural and physicochemical properties to ensure efficacious, bioavailability, and safety. They propose the concept of druggable proteins to bind potentially effective drug-like small molecules. One example is to identify metabolic enzymes as drug targets by

searching for similar structural properties of known drug targets in other organisms [57].

3.3.2. Coreness. A k -core of a graph G is a maximal connected subgraph of G in which all vertices have degree k at least. There is a way to determine a k -core by iteratively pruning nodes with a degree lower than k and their incident links [58]. The coreness of a protein is n if this protein is in the n -core of the PPI network but not in the $n + 1$ -core. We examined $rc_D(k)$ and $rc_{PT}(k)$ defined as follows for drug targets and other proteins in k -core, respectively.

$$rc_D(k) = \frac{p_k^D}{DR} \quad (5)$$

$$rc_{PT}(k) = \frac{p_k^{PT}}{PR},$$

where p_k^D and p_k^{PT} denote the proportion of the drug targets and other proteins in k -core of the PPI network, respectively. DR and PR are the drug target's ratio and pending test ratio in Table 1, respectively.

As $rd(i)$ is used in analysis of the modularity above, $rc_D(k) > 1$ indicates that the drug target proportion is higher than global level. In other words, the k -core tends to be a target-dominated subnetwork. In turn, if $rc_D(k) < 1$, the k -core mainly consists of nondrug target proteins. Figure 10 shows $rc_D(k)$ and $rc_{PT}(k)$ of the drug targets PPI networks.

Coreness poses a systematic way to consider the local and global significance of a protein, which indicates the inherent layer structure of the PPI network. Its biological significance has been found in several studies. For example, the functions of some function-unknown proteins of *E. coli* are predicted based on coreness [59]. The probability of yeast proteins both being essential and evolutionary conserved successively increases toward the innermost cores [60]. Disease-resistant domains have higher coreness than other and disease-susceptible domains [61]. In our study, the coreness shows significant difference between drug targets and normal proteins (see Figure 10). The drug target proteins are mainly in 6, 9, 12, 16, and 18 cores. In contrast, other proteins are evenly distributed over all the possible cores.

The proteins in the innermost k -cores, which are not necessarily among the highest connected ones, can interact with most high connection proteins [60]. As shown in Section 3.1, eigenvector centrality of the drug targets is not lower than normal proteins even though the known drug targets have lower degree and betweenness. For example, degrees of the drug targets PSMD1 and TOP2A are 33 which are less than normal proteins EWSR1 (135) and ATXN1 (171). But they are all in the 17 cores of the PPI network (see Figure 9). Moreover, the coreness of PSMD1 and TOP2A is 18 while coreness of EWSR1 (135) and ATXN1 is 17. Figure 9 shows the neighbors of the drug targets PSMD1 and TOP2A are less but have higher degree than normal proteins EWSR1 and ATXN1.

More generally speaking, coreness and eigenvector imply that the drug target proteins may not be the hub in the PPI network but they are able to propagate the affection to some

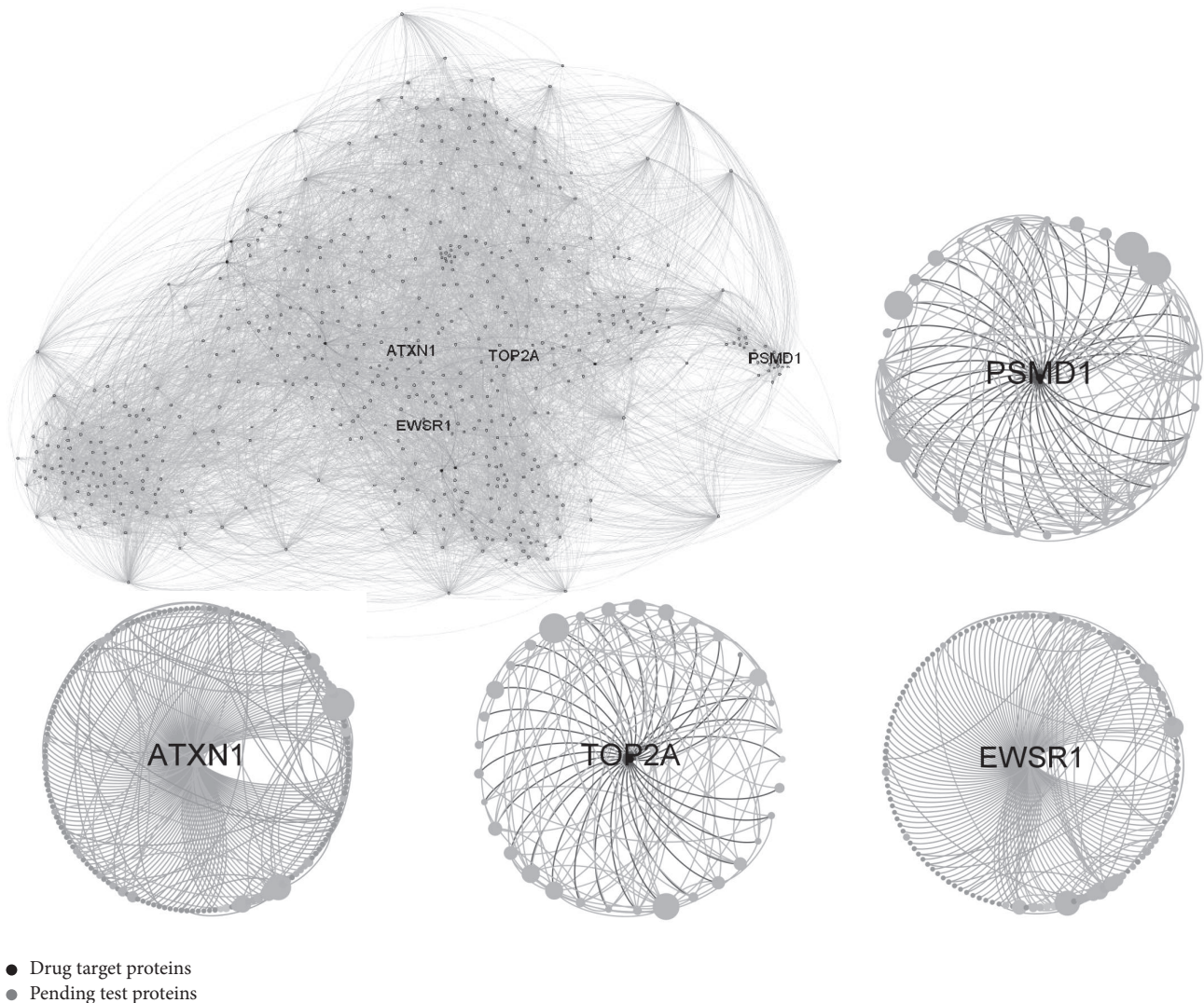


FIGURE 9: 17-core subnetwork of the drug targets PPI network. As shown in four ego networks of the known drug targets (PSMD1, TOP2A) and normal proteins (EWSR1, ATXN1), the neighbors of a drug target are less but those neighbors have higher degree than normal proteins. It implies a very important drug protein's reaction mechanism, that is, the drug target protein's interaction with most high connective proteins, though few of them are the hubs as important bridges of the PPI network.

hub-like proteins and spread the transcription signal to other related proteins through them. One of the possible reasons is that, in the cancer network, for example, only a part of the interactions among the related proteins may be active at a specific condition [62]. Therefore, the original hub is very likely to change into a normal protein. On the other hand, although the drug signal may be spread widely through hubs to inhibit the disease function, it is almost inevitable that it becomes an obstacle for many essential functions. But if drug stimulus is diffused indirectly, the drug signal may be more suitable for the tradeoff. According to the analysis about drug target structure characteristics, although the drug targets have few significant functions as intermediary and source of the drug, it is possible to identify potential drug targets based on their special structure characteristics.

4. Discussion

To understand the mechanism of drug targets at the molecular level where most drugs work, we studied the topological characteristics of the drug targets with three plausible views from which a drug target is known as (1) the intermediaries which play an important role on interactions of the drug targets network; (2) the sources which convert drug stimulus into the desirable therapeutic effects and spread to other proteins; (3) the proteins which have special topological and functional significance. From a series of their topological indices, we found somewhat surprising conclusions.

One conclusion was that known drug targets do not have the privilege of the first two roles although the drug fulfills its function by interactions with a target protein. In other

TABLE 2: The results of the experiments (10-fold cross-validation).

Algorithm	Using all features (8 topological features)			Using particular characteristics (3 topological features)		
	Accuracy	Positive predictive value	Negative predictive value	Accuracy	Positive predictive value	Negative predictive value
C4.5	65.2%	63.4%	66.5%	68.7%	67.7%	69.5%
Logistic regression	54.2%	52.8%	56.0%	67.6%	63.4%	68.0%
Naive Bayes	56.8%	54.4%	58.2%	67.3%	64.9%	68.0%
Bayes network	64.5%	63.4%	65.0%	70.6%	67.3%	71.5%
SVM	65.8%	60.2%	67.4%	72.0%	69.7%	73.4%
Radom forest	63.2%	60.4%	65.5%	69.6%	66.4%	70.5%

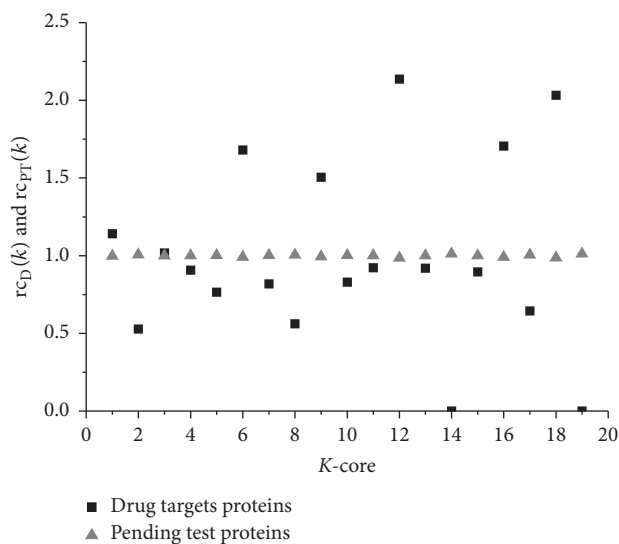


FIGURE 10: $rc_D(k)$ and $rc_{PT}(k)$ of the drug targets PPI networks. The drug target proteins are mainly in the 6-, 9-, 12-, 16-, and 18-core subnetworks, while the pending test proteins that represent normal proteins are evenly distributed across all core subnetworks.

words, the function of spreading drug stimulus seems not to be important for the drug targets. Meanwhile the drug targets have special topological structures that are different with normal proteins. We suspect that these special topological structures may help drug targets to respond to drug stimulus. Actually, the fact has been shown in some studies. For example, Overington et al. [4] suggested that most drugs depend on multiple specific motifs of the PPI networks. Hopkins [63] reported that it is necessary to map drug targets into integrated biological networks to identify the optimal points of protein-protein interactions for drug discovery. Ma'ayan et al. [64] argued that several classes of proteins with some special network statistics in the human genome appear to be better targets for drugs. Nacher and Schwartz [65] found that drugs usually have a high centrality value in the drug-therapy network and act on multiple molecular targets in the human system. Sakharkar et al. [66] showed that proteins with single-exon gene architecture are more likely to be targetable.

As the toy example shows the effectiveness of the features we discussed in the paper, we just use some traditional methods to detect drug targets with the features we discussed. These experiments concern two cases. One is to use all features including degree, betweenness, eigenvector centrality, average distance, eccentricity, modularity, clustering coefficient, community, and coreness, and the other is to use the drug target proteins' particular features. These three features were eccentricity, modularity, and coreness. This shows that the accuracy of the classifiers can benefit from using these particular features no matter what methods used to build classifier. All of the algorithms were performed on the original collected dataset with the help of Weka [67]. The details on results can be found in Table 2.

Although the causes are unknown currently, it is very helpful for drug discovery by reducing potential drug target proteins. Because at present the chemical and physical properties of known drug targets can be found but nondrug targets cannot, the prediction of drug targets is one classification problem for which there is no good solution. Currently, most studies usually use pending test proteins as nondrug targets, but they inevitably contain proteins that subsequently turn out to be drug targets although the majority of the proteins are normal proteins [28]. Therefore, there is an alternatively reasonable way to collect a more accurate nondrug targets dataset based on the drug targets structure characteristics.

Here we use modularity and coreness by naive Bayes to select nondrug targets from PT1 in which every protein consists of 39 chemical and physical properties. Assume that each protein has a 50% probability as a drug target in the pending test dataset. From 1,180 proteins in PT1, we selected $1,180 * 0.5 = 590$ proteins as the nondrug target dataset by the probabilities calculated by naive Bayes. By using the drug target dataset (D) and nondrug target dataset, we trained the drug target classifier based on support vector machines (SVMs) by Weka [67]. Finally, 102 proteins are predicted as potential drug target proteins from PT1. The accuracy of the classifier is 82.01% by 10-fold cross-validation. Positive predictive value is 72.7% and negative predictive value is 82.4%. These results indicated that SVMs could be reasonable in the prediction of drug target proteins. Particularly, some of the prediction results were found as published new drug targets recently. For example, CTBPI is listed by Cancer Resource [68], and TASPI is also a drug target, which is used

to combat novel diseases whose candidate genes are targeted by HNF4 α splice variants in hepatocellular carcinomas [69], as well as PNKP and RAG2 [70].

5. Conclusion

Protein-protein interactions are a better way to understand the biological functions through systemic view, which have been widely examined in many research studies under the molecular level. In this paper, the contrastive analysis on the topology of PPI networks between drug target proteins and other proteins provided a systemic biological mechanism of drug target protein interactions.

We found that 5 topological indices (degree, betweenness, eigenvector centrality, average distance, and cluster coefficient) are quite similar between drug target proteins and other proteins in the PPI network. It implies that known drug target proteins do not have the privilege of being a drug effect intermediary and/or source. It is different with some traditional views that intuitively consider a drug target protein as a hub of the PPI network.

On the other hand, a drug target protein has its own particular characteristics on the other 3 topological features including eccentricity, modularity, and coreness. It implies a drug target protein may have the capability to interact with some hub proteins, which pass their biological stimulus to other related proteins. These topological characteristics are helpful to understand how the drug target proteins work and test new drug effects. For example, one of the usages is to identify the new drug target. At present, there is no nondrug target dataset like the drug target dataset that records chemical and physical properties. However, according to these topological characteristics, the nondrug target proteins can be gathered. Hence, prediction of drug target proteins becomes a classic data mining problem which can be solved by many supervised learning algorithms. In the paper, 102 potential drug target proteins are predicted by SVMs, some of these have been published recently. It is just a simple example. One of the further works is to improve the prediction by examining these chemical and physical properties combined with topology. On the other hand, although there are improved methods to reduce the protein-protein interactions false positive rate [32, 71], the protein-protein interactions network always has a fairly high false positive rate. New high-throughput data and PPI database updates are important in the future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Chinese National Key Program of Basic Research (31000591, 31000587, and 31171266).

References

- [1] M. A. Lindsay, "Target discovery," *Nature Reviews Drug Discovery*, vol. 2, no. 10, pp. 831–838, 2003.
- [2] Y. Yang, S. J. Adelstein, and A. I. Kassis, "Target discovery from data mining approaches," *Drug Discovery Today*, vol. 14, no. 3–4, pp. 147–154, 2009.
- [3] P. Imming, C. Sinning, and A. Meyer, "Drugs, their targets and the nature and number of drug targets," *Nature Reviews Drug Discovery*, vol. 5, no. 10, pp. 821–834, 2006.
- [4] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?" *Nature Reviews Drug Discovery*, vol. 5, no. 12, pp. 993–996, 2006.
- [5] M. Rask-Andersen, M. S. Almén, and H. B. Schiöth, "Trends in the exploitation of novel drug targets," *Nature Reviews Drug Discovery*, vol. 10, no. 8, pp. 579–590, 2011.
- [6] M. A. Yildirim, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug-target network," *Nature Biotechnology*, vol. 25, no. 10, pp. 1119–1126, 2007.
- [7] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet, "Relating protein pharmacology by ligand chemistry," *Nature Biotechnology*, vol. 25, no. 2, pp. 197–206, 2007.
- [8] M. Campillos, M. Kuhn, A.-C. Gavin, L. J. Jensen, and P. Bork, "Drug target identification using side-effect similarity," *Science*, vol. 321, no. 5886, pp. 263–266, 2008.
- [9] J. De Las Rivas, C. Fontanillo, and F. Lewitter, "Protein-protein interactions essentials: key concepts to building and analyzing interactome networks," *PLoS Computational Biology*, vol. 6, no. 6, Article ID e1000807, 2010.
- [10] J. Bultinck, S. Lievens, and J. Tavernier, "Protein-protein interactions: network analysis and applications in drug discovery," *Current Pharmaceutical Design*, vol. 18, no. 30, pp. 4619–4629, 2012.
- [11] M. Zhu, L. Gao, X. Li et al., "The analysis of the drug-targets based on the topological properties in the human protein-protein interaction network," *Journal of Drug Targeting*, vol. 17, no. 7, pp. 524–532, 2009.
- [12] L. C. Royer, M. Reimann, B. Andreopoulos, and M. Schroeder, "Unraveling protein networks with power graph analysis," *PLoS Computational Biology*, vol. 4, no. 7, Article ID e1000108, 17 pages, 2008.
- [13] E. Chautard, N. Thierry-Mieg, and S. Ricard-Blum, "Interaction networks: from protein functions to drug discovery. A review," *Pathologie Biologie*, vol. 57, no. 4, pp. 324–333, 2009.
- [14] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.
- [15] D. Eisenberg, E. M. Marcotte, I. Xenarios, and T. O. Yeates, "Protein function the post-genomic era," *Nature*, vol. 405, no. 6788, pp. 823–826, 2000.
- [16] T. S. Keshava Prasad, R. Goel, K. Kandasamy et al., "Human protein reference database—2009 update," *Nucleic Acids Research*, vol. 37, no. 1, pp. D767–D772, 2009.
- [17] S. Kerrien, B. Aranda, L. Breuza et al., "The IntAct molecular interaction database in 2012," *Nucleic Acids Research*, vol. 40, no. 1, pp. D841–D846, 2012.
- [18] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "BioGRID: a general repository for interaction datasets," *Nucleic Acids Research*, vol. 34, supplement 1, pp. D535–D539, 2006.

- [19] L. Licata, L. Briganti, D. Peluso et al., "MINT, the molecular interaction database: 2012 update," *Nucleic Acids Research*, vol. 40, no. 1, pp. D857–D861, 2012.
- [20] L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, and D. Eisenberg, "The database of interacting proteins: 2004 update," *Nucleic Acids Research*, vol. 32, pp. D449–D451, 2004.
- [21] C. Knox, V. Law, T. Jewison et al., "DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs," *Nucleic Acids Research*, vol. 39, no. 1, pp. D1035–D1041, 2011.
- [22] G. Wang and R. L. Dunbrack Jr., "PISCES: a protein sequence culling server," *Bioinformatics*, vol. 19, no. 12, pp. 1589–1591, 2003.
- [23] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: signalP 3.0," *Journal of Molecular Biology*, vol. 340, no. 4, pp. 783–795, 2004.
- [24] A. Krogh, B. Larsson, G. Von Heijne, and E. L. L. Sonnhammer, "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes," *Journal of Molecular Biology*, vol. 305, no. 3, pp. 567–580, 2001.
- [25] J. C. Wootton and S. Federhen, "Statistics of local complexity in amino acid sequences and sequence databases," *Computers & Chemistry*, vol. 17, no. 2, pp. 149–163, 1993.
- [26] L. J. Jensen, R. Gupta, H.-H. Stærfeldt, and S. Brunak, "Prediction of human protein function according to Gene Ontology categories," *Bioinformatics*, vol. 19, no. 5, pp. 635–642, 2003.
- [27] K. Julenius, A. Mølgaard, R. Gupta, and S. Brunak, "Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites," *Glycobiology*, vol. 15, no. 2, pp. 153–164, 2005.
- [28] Q. Wang, J. Huang, Y. Feng, and J. Fei, "Efficient data mining algorithms for screening potential proteins of drug target," *Mathematical Problems in Engineering*, vol. 2017, Article ID 9852063, 10 pages, 2017.
- [29] P. Rice, L. Longden, and A. Bleasby, "EMBOSS: the European molecular biology open software suite," *Trends in Genetics*, vol. 16, no. 6, pp. 276–277, 2000.
- [30] T. M. Bakheet and A. J. Doig, "Properties and identification of human protein drug targets," *Bioinformatics*, vol. 25, no. 4, pp. 451–457, 2009.
- [31] J. Rual, K. Venkatesan, T. Hao et al., "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, no. 7062, pp. 1173–1178, 2005.
- [32] M. Barrios-Rodiles, K. R. Brown, B. Ozdamar et al., "High-throughput mapping of a dynamic signaling network in mammalian cells," *Science*, vol. 307, no. 5715, pp. 1621–1625, 2005.
- [33] Q. Li and L. Lai, "Prediction of potential drug targets based on simple sequence properties," *BMC Bioinformatics*, vol. 8, article no. 353, 2007.
- [34] F. Sams-Dodd, "Target-based drug discovery: is something wrong?" *Drug Discovery Today*, vol. 10, no. 2, pp. 139–147, 2005.
- [35] A. Schrattenholz and V. Šoškić, "What does systems biology mean for drug development?" *Current Medicinal Chemistry*, vol. 15, no. 15, pp. 1520–1528, 2008.
- [36] R. R. Vallabhajosyula, D. Chakravarti, S. Lutfeali, A. Ray, and A. Raval, "Identifying Hubs in protein interaction networks," *PLoS ONE*, vol. 4, no. 4, Article ID e5344, 2009.
- [37] E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, and A. Zinovyev, *Computational Systems Biology of Cancer*, CRC Press LLC, Boca Raton, Fla, USA, 2013.
- [38] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang, "High-betweenness proteins in the yeast protein interaction network," *Journal of Biomedicine & Biotechnology*, vol. 2005, no. 2, pp. 96–103, 2005.
- [39] 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 Computational Biology*, vol. 3, no. 4, pp. 713–720, 2007.
- [40] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, vol. 23, no. 3, pp. 191–201, 2001.
- [41] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology*, vol. 2, pp. 113–120, 1972.
- [42] G. Kar, A. Gursoy, and O. Keskin, "Human cancer protein-protein interaction network: a structural perspective," *PLoS Computational Biology*, vol. 5, no. 12, Article ID e1000601, 2009.
- [43] A. Özgür, T. Vu, G. Erkan, and D. R. Radev, "Identifying gene-disease associations using centrality on a literature mined gene-interaction network," *Bioinformatics*, vol. 24, no. 13, pp. i277–i285, 2008.
- [44] A. Gursoy, O. Keskin, and R. Nussinov, "Topological properties of protein interaction networks from a structural perspective," *Biochemical Society Transactions*, vol. 36, no. 6, pp. 1398–1403, 2008.
- [45] A. Mitsos, I. N. Melas, P. Siminelakis, A. D. Chairakaki, J. Saez-Rodriguez, and L. G. Alexopoulos, "Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data," *PLoS Computational Biology*, vol. 5, no. 12, Article ID e1000591, 2009.
- [46] K. A. Janes, J. G. Albeck, L. X. Peng, P. K. Sorger, D. A. Lauffenburger, and M. B. Yaffe, "A high-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: application to sepsis-apoptosis," *Molecular & Cellular Proteomics: MCP*, vol. 2, no. 7, pp. 463–473, 2003.
- [47] G. Scardoni, M. Petterlini, and C. Laudanna, "Analyzing biological network parameters with CentiScaPe," *Bioinformatics*, vol. 25, no. 21, pp. 2857–2859, 2009.
- [48] S. Wuchty and P. F. Stadler, "Centers of complex networks," *Journal of Theoretical Biology*, vol. 223, no. 1, pp. 45–53, 2003.
- [49] P. Hage and F. Harary, "Eccentricity and centrality in networks," *Social Networks*, vol. 17, no. 1, pp. 57–63, 1995.
- [50] R. C. Scott, D. Crabbe, B. Krynska, R. Ansari, and M. F. Kiani, "Aiming for the heart: targeted delivery of drugs to diseased cardiac tissue," *Expert Opinion on Drug Delivery*, vol. 5, no. 4, pp. 459–470, 2008.
- [51] R. A. Raj, "A recent technology in drug discovery and development," *International Journal of Innovative Drug Discovery*, vol. 2, no. 1, pp. 22–39, 2012.
- [52] R. A. E. Carr, M. Congreve, C. W. Murray, and D. C. Rees, "Fragment-based lead discovery: leads by design," *Drug Discovery Today*, vol. 10, no. 14, pp. 987–992, 2005.
- [53] J.-D. J. Han, N. Bertin, T. Hao et al., "Evidence for dynamically organized modularity in the yeast protein—protein interaction network," *Nature*, vol. 430, no. 6995, pp. 88–93, 2004.
- [54] I. W. Taylor, R. Linding, D. Warde-Farley et al., "Dynamic modularity in protein interaction networks predicts breast cancer outcome," *Nature Biotechnology*, vol. 27, no. 2, pp. 199–204, 2009.
- [55] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006.

- [56] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, Article ID P10008, 2008.
- [57] H. K. Majumder, *Drug Targets in Kinetoplastid Parasites*, Springer, Berlin, Germany, 2008.
- [58] V. Batagelj and M. Zaversnik, "An $O(m)$ algorithm for cores decomposition of networks," 2003, <https://arxiv.org/abs/cs/0310049>.
- [59] H. Hu, "F-curve, a graphical representation of protein sequences for similarity analysis based on physicochemical properties of amino acids," *MATCH. Communications in Mathematical and in Computer Chemistry*, vol. 73, no. 3, pp. 749–764, 2015.
- [60] S. Wuchty and E. Almaas, "Peeling the yeast protein network," *Proteomics*, vol. 5, no. 2, pp. 444–449, 2005.
- [61] C. M. Yates and M. J. E. Sternberg, "Proteins and domains vary in their tolerance of non-synonymous single nucleotide polymorphisms (nsSNPs)," *Journal of Molecular Biology*, vol. 425, no. 8, pp. 1274–1286, 2013.
- [62] A. Bossi and B. Lehner, "Tissue specificity and the human protein interaction network," *Molecular Systems Biology*, vol. 5, article 260, 2009.
- [63] A. L. Hopkins, "Network pharmacology: the next paradigm in drug discovery," *Nature Chemical Biology*, vol. 4, no. 11, pp. 682–690, 2008.
- [64] A. Ma'ayan, S. L. Jenkins, J. Goldfarb, and R. Iyengar, "Network analysis of FDA approved drugs and their targets," *Mount Sinai Journal of Medicine*, vol. 74, no. 1, pp. 27–32, 2007.
- [65] J. C. Nacher and J.-M. Schwartz, "A global view of drug-therapy interactions," *BMC Pharmacology*, vol. 8, article 5, 2008.
- [66] M. K. Sakharkar, P. Li, Z. Zhong, and K. R. Sakharkar, "Quantitative analysis on the characteristics of targets with FDA approved drugs," *International Journal of Biological Sciences*, vol. 4, no. 1, pp. 15–22, 2008.
- [67] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [68] M. D. Halling-Brown, K. C. Bulusu, M. Patel, J. E. Tym, and B. Al-Lazikani, "canSAR: an integrated cancer public translational research and drug discovery resource," *Nucleic Acids Research*, vol. 40, no. 1, pp. D947–D956, 2012.
- [69] M. Niehof and J. Borlak, "EPS15R, TASP1, and PRPF3 are novel disease candidate genes targeted by HNF4 α splice variants in hepatocellular carcinomas," *Gastroenterology*, vol. 134, no. 4, pp. 1191–1202, 2008.
- [70] C. A. Musselman and T. G. Kutateladze, "PHD fingers epigenetic effectors and potential drug targets," *Molecular Interventions*, vol. 9, no. 6, pp. 314–323, 2009.
- [71] R. Jansen, H. Yu, D. Greenbaum et al., "A Bayesian networks approach for predicting protein-protein interactions from genomic data," *Science*, vol. 302, no. 5644, pp. 449–453, 2003.