Identification and Evolution of Structurally Dominant Nodes in Protein-Protein Interaction Networks

Pei Wang, Xinghuo Yu, *Fellow, IEEE*, and Jinhu Lü, *Fellow, IEEE*

***Abstract*—It is well known that protein-protein interaction (PPI) networks are typical evolving complex networks. Identification of important nodes has been an emerging popular topic in complex networks. Many indexes have been proposed to measure the importance of nodes in complex networks, such as degree, closeness, betweenness, k-shell, clustering coefficient, semi-local centrality, eigenvector centrality. Based on multivariate statistical analysis, through integrating the above indexes and further considering the appearances of nodes in network motifs, this paper aims at developing a new measure to characterize the structurally dominant proteins (SDP) in PPI networks. Moreover, we will further investigate the evolution of the defined dominant nodes in temporal evolving real-world and artificial PPI networks. Our results indicate that the constructed artificial networks have some similar statistical properties as those of the real-world evolving networks. In this case, the artificial PPI networks can be used to further investigate the above evolution characteristics of the real-world evolving networks. Simulation results reveal that SDP in the yeast PPI networks are evolutionary conserved, however, the undominant nodes evolve rapidly. Furthermore, PPI networks are very robust against random mutations, while fragile yet with certain robustness to targeted mutations on SDP. Our investigations shed some light on the future applications of the evolving characteristics of bio-molecular networks, such as reengineering of particular networks for technological, synthetic or pharmacological purposes.**

***Index Terms*—Duplication-divergence model, evolution, important node, protein-protein interaction network, structurally dominant protein.**

I. Introduction

IT is well known that many real-world systems can be described by complex networks, such as social systems, biological systems, technological systems. Complex networks can be described by a graph $G = (V, E)$, and the graph G consist of $n = \|V\|$ nodes in node set V and $e = \|E\|$ edges in edge set E. Identifying important nodes in complex networks has been a hot topic in recent years [1]–[17]. In social networks, important nodes are also well-known as influential nodes, influential spreaders [9], [15], [16] or leaders [11]. Up to now, many indexes have been proposed to measure the importance of nodes in complex networks, such as degree, betweenness [2], clustering coefficient [4], closeness [4], [5], k-shell [6], eigenvector centrality [4], [7] and principal component centrality [8], semi-local centrality [9], PageRank [10] and LeaderRank [11], network motif centrality [17].

Degree $C_1(v)$ of node v counts the number of edges connected to v. It is simple yet intuitive, where the most connected nodes are the most important. Betweenness quantifies the number of times a node acts as a bridge along the shortest path between two other nodes [1], [2], which is defined for node v as

$$C_2(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

where σ_{st} is the total number of the shortest paths between node s and t, $\sigma_{st}(v)$ is the number of those paths that pass through v. It can well predict the possible flow that a node takes. Closeness $C_3(v)$ of node v measures how long it will spread information from node v to other connected nodes [5], [9], which is defined as

$$C_3(v) = \frac{1}{\sum\limits_{t \in V \setminus v} d_G(v, t)} \tag{2}$$

where $d_G(v, t)$ is the geodesic distance between nodes v and t in network G. The clustering coefficient $C_4(v)$ of node v is defined as the ratio of the number of triangles connected to node v to the number of triples centered on node v [3], [4]. The clustering coefficient for the whole network is defined as

$$CC = \frac{1}{n} \sum_i C_4(i). \tag{3}$$

The CC describes to what degree two nodes that are network neighbors of the same other nodes will themselves be neighbors [4]. K-shell $C_5(v)$ of node v can be obtained by recursively peeling all the nodes of degree less than k, until all nodes in the remaining network have degree at least k [6]. It can well quantify the global location of node v in the network. Eigenvector centrality $C_6(v)$ of node v is defined as the v'th component of the eigenvector corresponding to the maximal eigenvalue of the adjacency matrix of the network [4]. The recently proposed

Manuscript received November 28, 2013; revised January 05, 2014; accepted January 25, 2014. Date of publication March 04, 2014; date of current version March 25, 2014. This work was supported by the National Natural Science and Technology Major Project of China under Grant 2014ZX10004-001-014; the 973 Project under Grant 2014CB845302; the National Natural Science Foundation of China under Grants 61025017, 61304151 and11105040; the Australia ARC Discovery Grant DP130104765; and the Science Foundation of Henan University under Grants 2012YBZR007 and 2013YBRW005. This paper was recommended by Associate Editor S. T. C. Wong.

P. Wang is with the School of Mathematics and Information Sciences, Henan University, Kaifeng 475004, China, and also with the School of Electrical and Computer Engineering, RMIT University, Melbourne VIC 3001, Australia (e-mail: wp0307@126.com).

X. Yu is with the School of Electrical and Computer Engineering, RMIT University, Melbourne VIC 3001, Australia (e-mail: x.yu@rmit.edu.au).

J. Lü is with the Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China (e-mail: jhlu@iss.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBCAS.2014.2303160

semi-local centrality $C_7(v)$ of node v is defined as the number of the nearest and the next nearest neighbors of node v, which is proved to be more effective than the degree and betweenness in identifying influential spreaders in social networks [9]. The measures described above are typical indexes for complex networks, and have been frequently used to analyze the structure and function, as well as node importance of complex social networks [1]–[16].

However, compared with social networks, far less results on quantification of node importance have been reported for bio-molecular networks [14], [17]. Identification of important nodes in bio-molecular networks has crucial real-world implications. For example, in PPI networks or gene regulatory networks, through identifying important nodes, one can find disease related proteins or genes for pharmacological purposes [18]–[20]. Due to natural selection, bio-molecular networks evolve with time [21]. However, few results on identification of important nodes have been reported for evolving bio-molecular networks. The existing works have reported that a single measure in PPI networks can not identify lethal proteins from the others [22]. Thus, it is intriguing and fundamentally important to propose new integrative measures to explore important nodes in evolving bio-molecular networks and clarify their evolving characteristics.

Motivated by the above mentioned problems, we will propose a new integrative measure,which will integrate some of the mentioned well-known measures as an integrative index. These measures include degree, betweenness, closeness, clustering coefficient, k-shell, eigenvector centrality, semi-local centrality and the network motif centrality [17]. To distinguish the important node identified by the new integrative measure from the other definitions, we call the identified important nodes as structurally dominant proteins (SDP) or structurally dominant nodes (SDN), since the eight measures for these proteins tend to be high, therefore dominate specific functions of the PPI networks. Based on the well-defined new measure and the real-world yeast PPI networks, we will identify important nodes in evolving PPI networks, where networks will be constructed in Section II, the proposed new measure will be presented in Section III. Moreover, SDP in the PPI networks will be identified and evolving characteristics of such proteins will be clarified. Section IV investigates the robustness of the PPI networks against random and targeted mutations. Finally, some concluding remarks will be given in Section V.

II. Evolving Real-World and Artificial PPI Networks

A. Real-World PPI Networks

With the development of experimental technology for measuring interactions among proteins, such as the yeast two-hybrid screening (Y2H), the high-throughput affinity purification/mass spectrometry (AP/MS) and the yellow fluorescent protein complementation assay [23]–[25], a great quantity of data for PPIs of many organisms have been collected, and there have been many available online databases, such as BioGRID [26], YPD [27], MIPS [28], DIP [29]. In this paper, we consider the PPI networks for the Yeast Saccharomyces cerevisiae. We construct four networks with network size $n = 907, 1640, 1825$ and 2675, respectively. The first three networks are constructed by following the work of Schwikowski *et al.* [23]. The fourth network is based on the third one, and further includes 850 new proteins and 1933 interactions, which is reported by Yu *et al.* [24]. It is noted that the first network is based on the MIPS database, the second network further considers 917 proteins and 808 interactions (some proteins and interactions are the same as the first network) reported by Uetz *et al.* [30]. The third network is reported by Schwikowski *et al.* [23], which is based on the second one and further includes 185 proteins and their interactions from Ito *et al.* [31] and the DIP Database [29]. The obtained four original networks are all disconnected, their largest connected components contain 508, 1130, 1297 and 2307 nodes, and 796, 1619, 1862 and 3957 undirected edges, respectively (self-loops have been deleted). In this paper, we mainly investigate the largest connected components. The four connected networks consist of the evolving real-world PPI networks, with the former ones as subgraphs of the later ones.

TABLE I
Statistical Characteristics of Four Real-World (Above) and Four Artificial Networks (Below)

n	$\langle k\rangle$	k_{max}, k_{min}	APL	CC	PCC	PLE
508	3.1339	28,1	8.2249	0.1867	-0.0957	-2.1100
1130	2.8655	29,1	7.7299	0.1254	-0.0538	-2.6060
1297	2.8712	29,1	7.6270	0.1205	-0.0262	-2.6720
2307	3.4304	89,1	5.6359	0.0935	-0.0785	-2.5630
508	2.7795	45,1	6.1872	0.1688	-0.3615	-2.2340
1130	2.9646	68,1	5.8609	0.1745	-0.2804	-2.4580
1297	2.9915	72,1	5.7901	0.1680	-0.2677	-2.4620
2307	3.0611	106,1	5.6218	0.1263	-0.2065	-2.4470

Statistical characteristics of the four yeast networks are shown in the first four rows of Table I, where $\langle k\rangle, k_{\max}, k_{\min}$ represent average degree, maximum and minimum node degree, respectively, APL denotes average path length [4], CC denotes average clustering coefficient [4] as defined in (3), PCC represents Pearson correlation coefficient, which can measure the disassortativity of the network [4], [32]–[37], PLE denotes the power-law exponent [3], [4], [16]. From Table I, the average degrees of the four networks are all around 3, which indicate the sparsity of the networks. The gaps between the largest degree and smallest degree are very high, which provide clues for structural hierarchy. Moreover, the four networks have average path length range from 5.6 to 8.2, and clustering coefficients range from 0.09 to 0.18. It is noted that, for random networks, clustering coefficient can be approximated by $\langle k\rangle/(n-1)$ [38], obviously, the cluster coefficients for real-world PPI networks are larger than their random counterparts. Therefore, APL and CC values indicate the small world property of these networks. Negative Pearson correlation coefficients indicate the disassortativity [4], [37]. Finally, the degree distributions of the four PPI networks are all power-law, as shown in Fig. 1 (color on line, the red circles and downward sloping lines), the power-law exponents for the four networks are all around -2.5, which is consistent with the existing results in [4], [30]–[40], where PLE for biological networks are reported to be between -2 and -3.

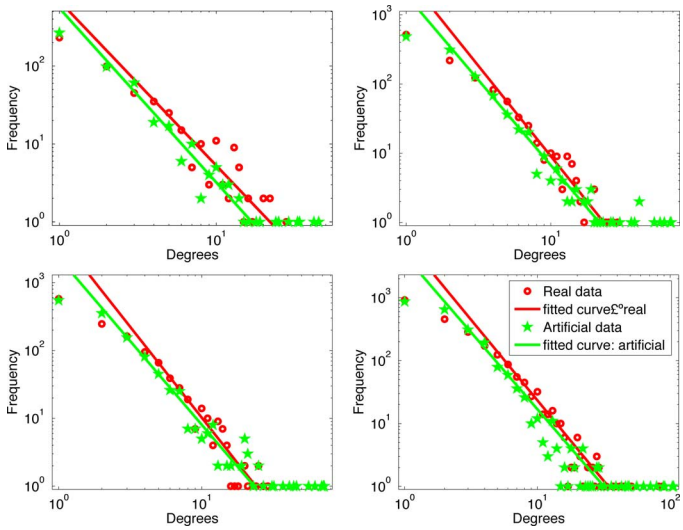

Fig. 1. Degree distribution of both real-world and artificial PPI networks. Network sizes are 508, 1130, 1297 and 2307.

B. Artificial PPI Networks

Duplication and divergence are two fundamental mechanisms for natural selection and evolution of biological networks. Based on the two mechanisms and other statistical characteristics of biological networks, such as sparsity, power-law degree distribution, small world effect and disassortativity, many artificial bio-molecular network generation algorithms have been proposed in the last decades [32]–[37]. For example, in 2002 and 2003, Solé *et al.* [32] and Vázquez *et al.* [33] firstly proposed the random DD models to generate artificial PPI networks. In 2003, Pastor-Satorras *et al.* [40] theoretically analyzed the evolving DD model. In 2007, Zhao *et al.* [37], investigated a random duplication model and an anti-preference duplication model to clarify the mechanism of disassortativity. In 2010, Wan *et al.* proposed a new DD algorithm to investigate the modularity and disassortativity of PPI networks [35]. Also in 2010, Xu *et al.* investigated how divergence affect the disassortativity of PPI networks [36]. Generally speaking, duplication and divergence can be reflected by various ways, different ways of duplication and divergence bring about different evolution models. For example, in network generation algorithms, duplication can be realized through random or anti-preference selection of targeted duplicated nodes. Divergence can be embodied by node deletion, edge deletion, edge addition, or edge rewiring and so on [36].

In the following, we use the anti-preference strategy to duplicate nodes. We randomly remove and add edges to introduce the divergence. Procedures of the algorithm are as follows [32]–[37]:

Step 1: Generate an initial connected network with n_0 nodes.

Step 2: Duplication. At each time step, node i with degree k_i is chosen with probability

$$p_i = \frac{\frac{1}{k_i}}{\Sigma_j \left(\frac{1}{k_j}\right)} \tag{4}$$

to duplicate. The replica i' has the same neighbors as node i.

Step 3: Divergence. Randomly choose a neighbor l of node i or its replica i', remove the undirected link (l, i) or (l, i') with probability α. Additionally, nodes i and i' can dimerize with probability β_0, that is, one can add an edge between them with probability β_0. $\beta_0 = \beta k_{i'}$, that is, it is proportional to the neighbours of node i' [35]. Furthermore, randomly choose a node $j(j \neq i)$, add a link between node j and i' with probability γ. After all these processes, remove isolated nodes.

Step 4: Stop when network reaches an expected size n.

We suppose $n_0 = 2$, that is, the network evolves from two ancestral nodes. For random duplication, from the theoretical mean-field analysis by Pastor-Satorras *et al.* [40], the best parameter α is $\alpha = 0.562$, γ satisfies

$$\gamma = \frac{(\alpha - 0.5)\langle k \rangle}{n} \tag{5}$$

where $\langle k \rangle$ is the average degree, n is the network size. But Pastor-Satorras *et al.* did not separate the dimerization processes from edge addition processes. Since dimerization between a duplicated node and its replica can happen with a greater probability [35], β should be greater than γ. Furthermore, we use the anti-preference duplication mechanism, which can strengthen the disassortativity [37]. In the following, we set $\alpha = 0.562$, $\beta = 0.1$, $\gamma = 0.000165$ for all the simulation runs. To keep the evolving structure of the artificial networks, when we generate networks with larger sizes, we take the newly generated smaller network as the initial network. Statistical characteristics of four artificial PPI networks are presented in Table I, with expected network size $n = 508, 1130, 1297, 2307$, respectively. From Table I, the average node degrees for the artificial networks range from 2.7795 to 3.0611, which are in good agreement with the real-world ones. The APL are all around 6 and the clustering coefficients are all around 0.16, which are also consistent with the real-world networks, and indicate the small world property [16]. The PCC values are all negative, which indicates the disassortativity [37]. The absolute values of PCC are a little larger than that of the real-world networks. The degree distribution of the artificial networks are also power-law, with exponents all approximate to -2.5. The degree distributions of the artificial networks with fitted curves are shown in Fig. 1 (Color online, green stars and downward sloping lines). As one can see from Fig. 1, the degree distributions for the artificial networks are all consistent with the real-world ones. To summarize, by comparing the statistical indexes between the real-world and artificial networks, we find the artificial networks are in good agreement with the real-world ones. Furthermore, for the four artificial PPI networks, with the evolving of network sizes, the average degree and maximum degree increase. The average path length, the clustering coefficients and the absolute values of PCC decrease with the increasing of network size. The evolutionary trends of these indexes are roughly the same as the real-world networks, which indicate that the parameters we have chosen in the DD model are biologically relevant, and one can use artificial networks to investigate the evolution characteristics of real-world ones.

TABLE II
NETWORK MOTIFS IN THE REAL-WORLD (LEFT) AND ARTIFICIAL NETWORKS (RIGHT). $N_{r.}, Z_{s.}$ DENOTE N_{real}, Z_{score}

			Real.	nets		Arti.	nets	
n	Size	ID	$N_{r.}$	$Z_{s.}$	U	$N_{r.}$	$Z_{s.}$	U
508	3	238	667	105.99	39	141	5.81	30
	4	4958	2009	3.87	32			
		13278	120	10.03	16			
		31710	1381	2382.86	13	17	2.48	7
1130	3	238	776	141.39	62	359	9.14	62
	4	4958	3687	13.24	56			
		13278	218	38.00	28	639	2.42	29
		31710	1416	4922.79	18	47	4.04	13
1297	3	238	810	169.20	69	416	14.03	71
	4	4958	4200	20.17	63			
		13260	151	5.73	28			
		13278	274	78.84	32	722	2.55	34
		31710	1425	10127.41	19	58	4.60	15
2307	3	238	1072	66.62	116	560	11.19	118
	4	4958	18007	4.18	107			
		13260	3338	45.93	53			
		13278	727	12.31	53			
		31710	1457	921.66	27	53	3.30	18

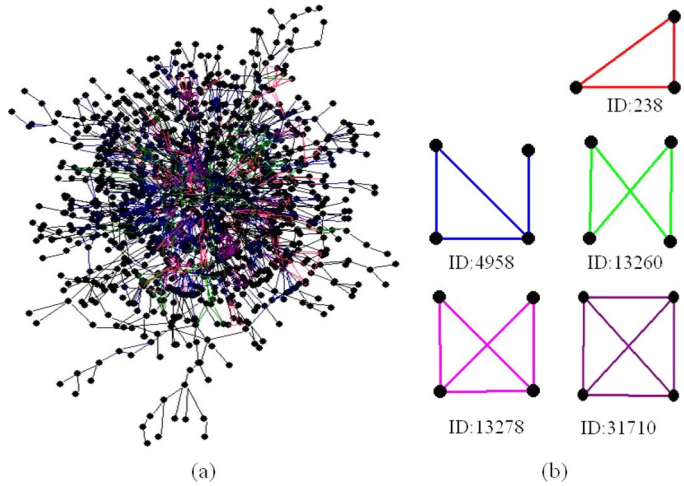

Fig. 2. The Yeast PPI Network with 1297 nodes and motifs. Edges involve in network motifs in network (a) are marked out by colors as shown in (b) (color online). (a) Real-world network with 1297 nodes. (b) 3-node and 4-node network motifs in (a).

C. Network Motifs in PPI Networks

The concept of network motif is firstly proposed by Milo and co-authors in 2002 [41], [42]. Network motifs are overrepresented subgraphs, which appear in a complex network more frequently than that in its random counterparts.

mDraw (http://www.weizmann.ac.il/mcb/UriAlon) is a user friendly visualization tool for networks and motif detection, which is developed by David Harel and colleagues. In the following, we use mDraw to detect network motifs in the constructed real-world and artificial PPI networks. The detected motifs in the networks are summarized in Table II, where, for each network, we have generated 100 random networks to compare with the investigated networks. In Table II, the column n denotes network size; the column $Size$ denotes motif size, where we have considered 3-node and 4-node motifs; the column ID denotes the identification number of a motif, which is defined by the adjacency matrix of the motif (see the maual for mDraw); the column N_{real} represents number of appearance of a subgraph in the investigated network; Z_{score} is defined as

$$Z_{score} = \frac{N_{real} - N_{rand}}{SD} \tag{6}$$

where N_{rand} denotes the average number of appearance of the subgraph in 100 random networks, SD denotes standard deviation of N_{rand}. Z_{score} can quantify the statistical significance of a subgraph [41]. U is defined as the number of times a subgraph appears in the investigated network with distinct sets of nodes [41]. In this paper, subgraphs with $Z_{score} \geq 2$ and $U \geq 4$ are taken as network motifs.

Fig. 2 shows the real-world yeast PPI networks with 1297 nodes, as well as the detected network motifs. From Table II and Fig. 2, we can see that both real-world and artificial PPI networks contain the 3-node motif 238 and the 4-node motif 13278 and 31710. Obviously, motifs in the real-world PPI networks are more abundant and more diverse than that in the artificial ones. In the artificial networks, subgraph 4958 and 13260 are not network motifs. However, the generated artificial networks can reflect the most of the characteristics of the real-world ones.

Why are subgraph 238 and 31710 and others the building blocks of PPI networks? From evolutionary perspective, duplication and divergence are two basic mechanisms for biological evolution. During DD processes, on the one hand, many square structures can be produced by the duplication processes, since the target node and its replica share the same neighbors. On the other hand, the dimerization processes tend to bridge the target node and its replica with a probability, which can produce the triangular structure and 4-node fully connected structure. Therefore, one can conclude that the cooperation between duplication and dimerization bring about the motif 238 and 31710, while random edge deletion and addition processes produce the other motifs.

III. STRUCTURALLY DOMINANT PROTEINS AND ITS EVOLUTION IN PPI NETWORKS

A. The New Integrative Measure of Node Importance

Integrating some existing measures of node importance, based on the multivariate statistical analysis method [43], [44], we will propose a new integrative measure of node importance in PPI networks. Our main idea is as follows. Firstly, we derive degree vector C_1, betweenness vector C_2, clossness vector C_3, clustering coefficient vector C_4, k-shell vector C_5, eigenvector centrality C_6, semi-local centrality vector C_7 for each network. Moreover, we count the appearances of each node in all network motifs and form another vector C_8, which can reflect the importance of each node in complex biological networks [17], since Milo *et al.* has suggested that network motifs are building blocks of complex biological networks [41], [42], and it has been evidenced that network motifs always play crucial functional roles in biological systems [45]–[49]. Secondly, based on the obtained $m = 8$ measures, we perform principal component analysis (PCA) [39], [43], [44] and derive the first

TABLE III
CONTRIBUTION OF THE FPC AND WEIGHT VECTORS FOR THE REAL-WORLD AND ARTIFICIAL PPI NETWORKS

Network	n	λ_1	ρ	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
Real-World	508	4.5108	56.38%	0.3744	0.1313	0.1342	0.2488	0.4374	0.4126	0.4459	0.4498
	1130	4.4887	56.11%	0.3730	0.1986	0.2310	0.2115	0.4290	0.3876	0.4479	0.4341
	1297	4.4977	56.22%	0.3750	0.2120	0.2399	0.2041	0.4266	0.3844	0.4481	0.4297
	2307	4.4383	55.48%	0.4162	0.3558	0.3107	0.0910	0.3521	0.3596	0.4027	0.4242
Artificial	508	3.6011	45.01%	0.4502	0.3964	0.3052	0.1410	0.3727	0.2779	0.4121	0.3758
	1130	3.7939	47.42%	0.4217	0.3837	0.3574	0.0660	0.3228	0.3950	0.4345	0.3064
	1297	3.7103	46.38%	0.4285	0.3967	0.3590	0.0668	0.2955	0.3920	0.4333	0.3110
	2307	3.7494	46.87%	0.4304	0.4192	0.3558	0.0684	0.2778	0.3703	0.4119	0.3518

principal component (FPC) as a new integrative measure, which reads as follows:

$$FPC = \sum_{i=1}^{m} w_i C_i. \tag{7}$$

The FPC is actually a weighted sum of the m measures, the weight $w = (w_1, w_2, \ldots, w_m)^T$ is an unit vector, which is to be determined. It is noted that the FPC is the extension of the general summation operation. The weighted sum FPC is much better than the traditional direct sum, since different components often have different importance for determining the specific functions of networks. The existing works have reported that hubs are important nodes [14], some suggest that nodes bridging different communities are important ones [5], while the other works have found nodes with the highest semi-local [9] or k-shell [15] are more influential. From the PCA theory, the first principle component can act as an ordering index. Therefore, we have extended the PCA theory to complex networks. The FPC integrates the eight given structural measures. The proteins with high FPC values are much more likely to be the structurally dominant nodes in the PPI networks. In this case, the proteins with high FPC values often play much more important roles than those of the proteins with low FPC values. Therefore, the proteins with high FPC values are dominant proteins for the functions of PPI networks. As we have discussed in the Introduction, we call such important nodes as SDP or SDN. The best index FPC should have the best discriminabiltity among nodes, therefore, the variance of FPC should be as large as possible. Thus, w can be determined by following the rule of making the variance of FPC to realize maximization [39], [43], [44]. Take $\mathcal{C}_1, \ldots, \mathcal{C}_m$ as random variables, which represent the m measures of a node in complex networks, and the n dimensional values $C_1, \ldots, C_m$ from a concrete network as its n observations, denote $\mathcal{C} = (\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m)^T$ and its observation matrix as $C = (C_1, C_2, \ldots, C_m)_{n\times m}$. The covariance matrix of $\mathcal{C}$ can be estimated by its observation C. Denote the covariance matrix of C as Σ, then

$$Var(\mathcal{C}) \approx Var(C) = \Sigma = \frac{1}{n-1}(C^T C - n\overline{C}\,\overline{C}^T) \tag{8}$$

where $\overline{C}$ is the column mean vector of C, n is network size. Indeed, Σ is the unbiased estimator of $Var(\mathcal{C})$. Based on the above notations, we have the stochastic form of FPC as $\mathcal{FPC} = w^T\mathcal{C}$ and

$$Var(\mathcal{FPC}) = Var(w^T\mathcal{C}) = w^T Var(\mathcal{C}) w \approx w^T \Sigma w. \tag{9}$$

The unit vector w is therefore determined by the following optimization problem:

$$\begin{aligned} \max \quad & w^T \Sigma w \\ s.t. \quad & w^T w = 1. \end{aligned} \tag{10}$$

By the Lagrangian multiplier method, let

$$L(w, \lambda) = w^T \Sigma w - \lambda(w^T w - 1) \tag{11}$$

the optimal w satisfies

$$\begin{cases} \frac{\partial L}{\partial w} = 2(\Sigma - \lambda I)w = 0, \\ \frac{\partial L}{\partial \lambda} = 1 - w^T w = 0 \end{cases} \tag{12}$$

where I is the identity matrix. The first sub-equation of (12) implies that $\Sigma w = \lambda w$, that is, λ and w are just the eigenvalue and eigenvector of Σ, respectively. Furthermore, we have $Var(\mathcal{FPC}) \approx w^T \Sigma w = \lambda w^T w = \lambda$. Therefore, the maximum value of $w^T \Sigma w$ equals the biggest eigenvalue of Σ, denoted λ^*, the optimal vector w^*, satisfying that $w^T \Sigma w$ reaches its maximum value, is just the corresponding unit eigenvector of λ^*.

The covariance matrix Σ is a real symmetric matrix, therefore, its eigenvalues are all nonnegative. Denote $\lambda_1 \geq \lambda_2 \geq \cdots, \geq \lambda_m \geq 0$ as the eigenvalues of Σ, then $\lambda^* = \lambda_1$. The index $\rho = \lambda_1 / \sum_{i=1}^{m} \lambda_i$ is the contribution ratio of the FPC, which can measure the amount of information in C extracted by the FPC [39], [43], [44].

Up to now, we have determined the integrative measure. Finally, for a concrete network, substituted concrete values of C_i in (7), one derives the observation of $\mathcal{FPC}$ as FPC. Each node has a FPC value, based on the FPC value, one can identify SDN in complex networks. Nodes with the largest FPC values are SDN. It is noted that, if node v has larger $C_i (i = 1, 2, \ldots, m)$ than other nodes, then the FPC value must be also larger than other nodes. The integrative index FPC considers various structural characteristics of each node in a network, and obviously more reasonable than a single index.

B. Identifying SDP in Evolving PPI Networks

For the eight networks in Section II, the maximum eigenvalue of Σ, the contribution of the FPC, as well as the weight vector are obtained in Table III. It is noted that values in matrix C have different orders of magnitudes, for example, the magnitude of the eigenvector centrality ranges from 10^{-1} to 10^{-13}, while the magnitude of the semi-local centrality ranges from 10 to 10^4. In order to avoid the malfunction of indexes with low order of

TABLE IV
THE IDENTIFIED TOP-10 RANKED NODES IN THE REAL-WORLD (LEFT) AND ARTIFICIAL (RIGHT) PPI NETWORKS AND THE CORRESPONDING RANKS BY C_i. ND AND PROTE ARE THE ABBREVIATION OF NODE AND PROTEIN, RESPECTIVELY

			Real. net.									Arti. net.								
n	Nd	Prote	R_F	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	Nd	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
	89	TAF10	1	6	53	149	84	4	1	2	1	15	1	2	31	152	16	1	2	5
	120	ADA2	2	8	118	150	69	7	2	3	3	2	2	1	1	156	9	54	1	14
	228	TRA1	3	13	18	25	65	10	7	1	5	14	3	4	143	150	15	10	20	11
	92	NGG1	4	7	86	153	83	5	6	7	2	22	4	5	2	155	20	82	7	27
508	214	GCN5	5	9	74	122	68	9	5	6	4	142	7	6	13	149	45	116	3	29
	114	SPT3	6	12	248	151	63	6	3	4	6	72	5	45	268	148	2	147	29	4
	451	SPT20	7	14	474	152	64	13	4	5	7	153	27	72	109	82	3	180	26	1
	27	SPT7	8	15	201	154	3	1	8	8	8	56	8	3	6	153	29	148	80	58
	41	TAF5	9	16	209	155	6	2	9	9	9	144	23	16	146	123	46	4	5	15
	87	TAF12	10	17	235	156	13	3	10	10	10	17	6	26	102	147	17	135	17	12
	121	TAF10	1	7	81	62	133	4	1	3	1	2	1	1	1	364	8	2	1	45
	158	ADA2	2	9	53	29	118	7	2	2	2	142	2	3	2	365	27	1	2	60
	125	NGG1	3	10	216	89	117	5	6	8	3	677	8	13	9	345	51	3	3	1
	518	TAF9	4	15	89	35	94	12	7	6	4	15	4	2	5	358	64	315	4	34
1130	304	TRA1	5	21	58	27	89	10	8	1	8	72	3	10	39	359	2	9	6	8
	287	GCN5	6	14	185	59	93	9	5	7	5	697	16	216	42	312	52	4	14	4
	233	TAF6	7	20	137	86	88	8	9	9	7	14	5	8	55	360	63	347	18	18
	151	SPT3	8	19	554	87	85	6	3	4	9	22	6	9	6	362	13	48	13	109
	591	SPT20	9	22	793	88	86	13	4	5	10	17	19	118	56	310	12	5	19	5
	119	TAF12	10	18	400	90	87	3	10	10	6	262	32	20	4	331	37	6	5	24
	127	TAF10	1	6	60	56	144	4	1	2	1	2	1	1	1	413	48	2	1	41
	164	ADA2	2	9	28	27	129	7	2	1	2	142	2	3	2	415	81	1	2	82
	131	NGG1	3	10	82	47	122	5	3	4	3	677	8	10	9	393	36	3	3	1
	542	TAF9	4	15	92	40	97	12	7	7	4	15	4	2	3	410	54	242	4	38
1297	315	TRA1	5	21	52	30	91	10	8	3	10	72	3	8	47	409	2	23	6	7
	298	GCN5	6	14	241	57	96	9	6	8	5	14	5	6	30	411	53	375	14	20
	157	SPT3	7	19	659	84	87	6	4	5	6	697	18	190	62	359	37	4	12	4
	619	SPT20	8	22	893	86	88	13	5	6	7	17	21	127	112	356	8	5	20	5
	243	TAF6	9	20	134	85	90	8	9	9	9	22	7	9	7	414	56	63	13	188
	125	TAF12	10	18	468	90	89	3	10	10	8	873	15	21	63	378	154	468	21	3
	1232	GCD7	1	1	1	1	540	61	15	1	1	2	1	1	1	636	42	1	1	6
	1252	ATG17	2	3	2	2	543	62	16	3	4	142	2	3	2	641	84	2	2	24
	662	SRP1	3	2	3	3	547	51	17	2	3	15	5	2	3	633	47	241	4	15
	189	ADA2	4	8	10	37	384	7	1	26	2	677	10	11	5	611	152	3	3	2
2307	151	NGG1	5	28	51	41	174	5	2	31	6	72	4	8	17	632	2	215	7	5
	147	TAF10	6	24	100	230	176	4	3	71	7	14	3	4	9	637	46	247	5	17
	200	SMT3	7	5	4	6	541	84	59	29	22	41	6	5	22	634	56	424	35	22
	622	TAF9	8	38	137	236	148	12	6	74	8	22	7	6	10	638	50	214	12	72
	1428	MUK1	9	17	34	8	470	67	20	6	5	150	13	13	20	623	87	8	11	39
	705	SPT20	10	51	270	246	122	13	4	69	9	13	16	7	13	619	8	290	19	46

magnitude, before our analysis, we have performed standardized transformation to each column of C.

From Table III, we see that the maximum eigenvalues of Σ for each real-world network are around 4.5, while for artificial networks, the values are around 3.7. The contribution of the FPC for each real-world network is all above 55%, while the contribution is a little lower for artificial ones. This indicates that the variance among the eight indexes for the real-world networks is larger than that for the artificial ones, and the FPC index for the real-world networks can have better performance than for the artificial ones. For the real-world networks, the values for $w_i(i = 5, \ldots, 8, 1)$ are relatively larger than $w_j(j = 2, 3, 4)$. This indicates that the k-shell, the eigenvector centrality, the semi-local centrality, the network motif centrality and the degree play more important roles than the other indexes in the FPC. For the artificial networks, except w_4, there are no much differences among the others. Low value of w_4 indicates the clustering coefficient index has very weak contribution to the FPC.

For the real-world networks, by the FPC, the top-10 ranked nodes and their corresponding ranks by the other indexes are shown in the left column of Table IV, where we have listed the corresponding protein names and their ranks R_i by the measure $C_i(i = 1, \ldots, 8)$. From Table IV, for each network, one can conclude that the rank by the integrative FPC index is different from all the eight single indexes. The $R_i(i = 5, \ldots, 8, 1)$ have the best consistence with the R_F from the the FPC measure, it indicates nodes with high $C_i(i = 5, \ldots, 8, 1)$ are more prone to be SDN. In Indeed, this result just corresponds to the result in Table III, where, the weight coefficients for $C_i(i = 5, \ldots, 8, 1)$ are higher than the other indexes. By comparing the results among different real-world networks, more than 50% of the top-10 proteins are almost the same in different networks. With the evolution of network size, some nodes transit from not the top-10 to the top-10 ranked nodes, and even become the highest ranked nodes, such as GCD7,ATG17,SRP1 and SMT3. Whereas, a few nodes become less and less important, such as TRA1, TAF12, SPT7 and TAF5.

From the Gene Ontology database [50], we can get more information about the functions of the proteins in the real-world networks. The twelve proteins:TAF10, ADA2, TRA1, NGG1, GCN5, SPT3, SPT20, SPT7, TAF5, TAF12, TAF9, TAF6 are all subunits of the SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex, which perform similar functions or involved in the same biological processes. The SAGA complex is important for transcription in vivo and possesses histone acetylation function, mu-

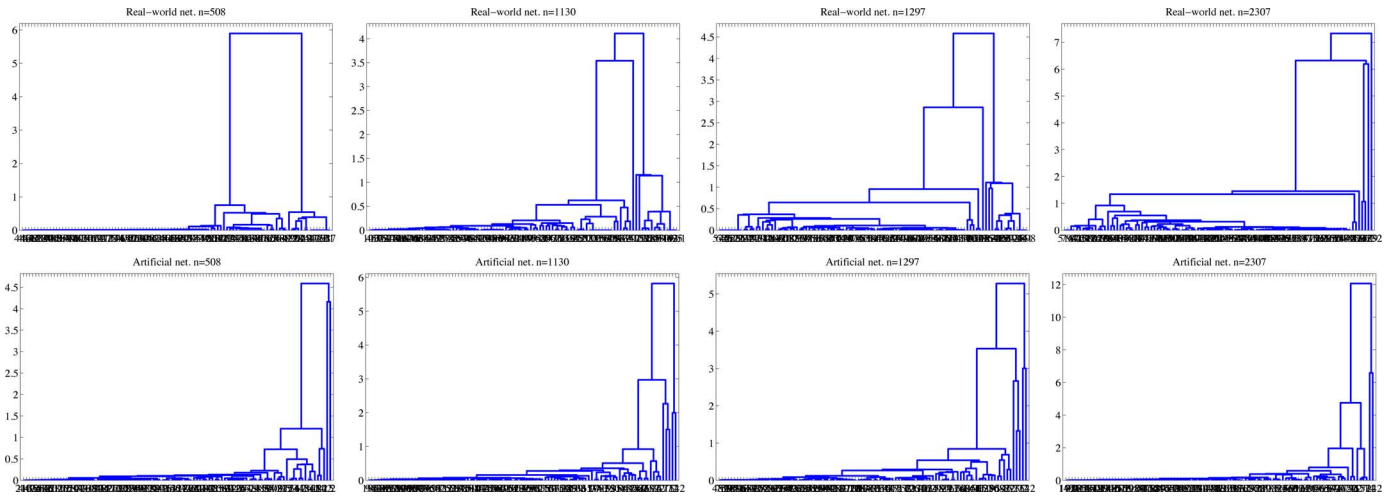

Fig. 3. Clustering dendrogram of the top-100 ranked nodes for the real-world and artificial networks.

tations in SPT7 or SPT20 will disrupt the SAGA complex and cause severe phenotypes [51]. Furthermore, the SAGA complex also has important roles in transcript elongation, the regulation of protein stability, and telomere maintenance [52]. From Bertolazzi *et al.* [14], protein complexes are often evolutionary conserved, here, with the evolution of networks, some of the SAGA complex can keep to be the top-10 ranked proteins, therefore, they are functionally important nodes in the yeast PPI networks. For the rest proteins in Table IV, the reduction of GCD7 function will lead to decrease of competitive fitness and decrease of resistance to methyl methanesulfonate [50]. Over-expression of ATG17 inhibits filamentous growth. Mutation of ATG17 leads to bud morphology abnormal, the decrease of competitive fitness and so on [50]. Reduction of SRP1, SMT3 and MUK1 functions all lead to the decrease of competitive fitness, over-expression of SRP1 can resort to invasive growth increased and vegetative growth rate decreased [50], and over-expression of SMT3 can lead to decreased vegetative growth. Over-expression of MUK1 resorts to the abnormal of vacuolar transport [50]. Obviously, most of the identified top-10 proteins are essential for the growth and reproduction of the Yeast.

For the artificial networks, the top-10 ranked nodes by the FPC and their corresponding ranks by the other indexes are obtained in the right column of Table IV, one can see that the ranks by the indexes with larger weight have better consistence with R_F. The clustering coefficient method is the most unusual, for each network, the clustering coefficients of the identified top-10 ranked nodes are all very small. However, by comparing the results among the different artificial networks, we can conclude that with the evolution of network size, most of the top-10 ranked nodes can maintain at the top-10 level.

Fig. 3 shows the clustering dendrogram of top-100 ranked nodes for each networks. From Fig. 3, we can see that both the real-world and the artificial PPI networks have clear hierarchical structures. Nodes can be classified into clusters, with the most important cluster on the right. The most important cluster only contains a few nodes. A large amount of nodes are not SDN. The phenomenon just indicates the scale-free structural feature of these networks. Fig. 4 displays the ranking-based rich-club

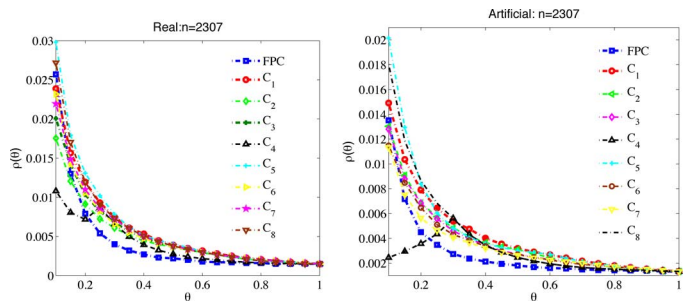

Fig. 4. Rich-club analysis by the FPC and the other ranking measures.

TABLE V
CORRELATION COEFFICIENTS BETWEEN FPC AND C_i

	Real. nets				Arti. nets			
C_i	508	1130	1297	2307	508	1130	1297	2307
C_1	.6364	.6340	.6314	.7848	.8219	.8101	.8130	.8242
C_2	.1315	.2870	.3193	.7484	.7634	.7748	.7914	.8528
C_3	.1069	.1906	.2019	.2887	.1795	.2452	.2392	.2451
C_4	.3188	.2724	.2559	.0458	-.048	-.079	-.069	-.054
C_5	.8410	.8055	.7882	.5186	.3796	.3567	.2752	.2157
C_6	.9209	.8839	.8810	.7584	.4457	.6854	.6798	.6687
C_7	.8885	.8771	.8834	.7433	.4983	.5893	.5746	.5628
C_8	.9516	.9410	.9355	.9257	.6111	.5536	.5622	.6262

phenomenon for the PPI networks with 2307 nodes, where $\rho(\theta)$ denotes the density of links among the top-$100\theta\%$ ranked nodes [9], [39]. In Fig. 4, we have shown the outcomes of the FPC ranking method and the other eight methods. The monotone decreasing of $\rho(\theta)$ with θ indicates rich-club ordering. By comparing between different methods, the clustering coefficient C_4 has the worst performance, the curves for the two networks are all not monotone decreasing, while the proposed FPC measure has very good performance, where top-ranked nodes are highly connected, and the curve from our method decreases faster than the other method.

Table V shows the correlation coefficients between FPC and C_i for all the networks. For real-world networks, the correlation coefficients between FPC and $C_i (i = 5, \ldots, 8, 1)$ are all very high, one can say that the FPC index is highly affected

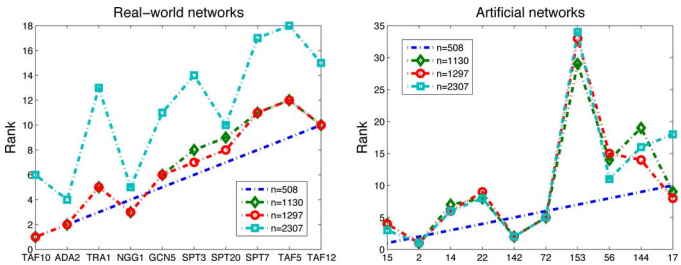

Fig. 5. Evolution of the top-10 SDP in networks with 508 nodes.

by these measures. Furthermore, the FPC index is the most relevant with the network motif centrality C_8, which results in the ranking similarity between the two measures. For artificial networks, the correlation coefficients between FPC and $C_i(i = 1, 2, 6, 7, 8)$ are larger than the others, which indicate nodes in artificial networks with larger degree, betweenness, eigenvector centrality, semi-local centrality or network motif centrality are more prone to be SDN. The correlation coefficient between FPC and C_4 are negative, which indicate the identified SDN tend to be with low closeness values. As we can see from Table IV, the R_4 of the identified top-10 ranked nodes are all very large.

C. Evolution of SDP in PPI Networks

In the following, we discuss the evolutionary characteristics of SDP in PPI networks. For simplicity, we only consider the evolution of the top-10 nodes in networks with 508 nodes. Fig. 5 shows the ranks of top-10 nodes in networks with 508 nodes and their ranks in larger networks. The line corresponding to $n = 508$ denotes the ranks in the network with size 508, and this line acts as a baseline. The ranks of the considered nodes in larger networks consist of another three curves. By comparing the cases between the real-world and artificial networks, similar phenomenon can be observed. The curves for larger networks are mostly above the baseline, and therefore, we can say for PPI networks, roughly speaking, the ranks of the most of the considered nodes drop with the increasing of network sizes. However, for the real-world networks, the ranks of protein ADA2, NGG1, SPT20 have no much changes as compared with other proteins. The observations tell us that SDP are evolving with the evolving of networks, some SDP keep to be dominant ones, while a large amount of nodes become less and less dominant with the evolution of the PPI networks.

From Section II and the above investigations, we have concluded that the real-world and artificial networks have similar statistical characteristics, therefore, we can generate extensive artificial networks with different structures and sizes to infer the evolutionary characteristics of the real-world ones. To further investigate the evolution of SDP, we randomly generate 5 sets of networks with sizes $n = 508, 800, 1130, 1297, 1700, 2307$, and further explore their evolutionary characteristics. By averaged over the simulation runs of these networks, Fig. 6 draws the evolution of ranks for the first 508 nodes in the first networks. The 508 nodes in the first network are sorted by their ranks in descending order, as described by the first column corresponding to $n = 508$. The rest columns show the ranks of the

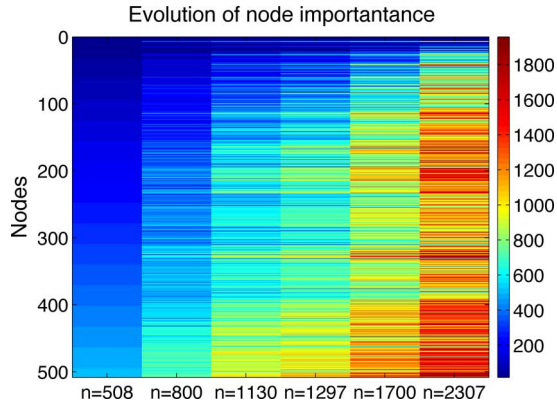

Fig. 6. Evolution of 508 nodes in artificial PPI networks. Averaged over 5 sets of networks.

508 nodes in larger networks. From Fig. 6, we can see that with the evolution of networks, only a few SDN in smaller networks can keep to be dominant in larger networks, while for a large amount of nodes, their ranks increase with network size, which indicate that they become less and less dominant. Furthermore, less dominant nodes in smaller networks can becomes less and less dominant in larger networks more rapidly, which indicates that SDN evolve slower than undominant ones. From the review paper of Jancura and Marchiori [53], a generally accepted premise for the evolution of proteins in networks is that essential proteins should evolve at slower rates than non-essential ones. Interestingly, since we have found SDP are prone to be essential ones, therefore our observations provide some evidences for the hypothesis.

IV. Robustness Against Mutations

In 2000, Barabasi *et al.* [54] found that scale-free complex networks have the property of random error tolerance and attack vulnerability. In 2001, it was found that PPI networks are lethality to the deletion of highly connected proteins [55]. Here, we investigate the robustness of PPI networks against targeted and random mutations. For targeted mutations, we consider both the removal of highly connected nodes and nodes with high FPC values. The curves of the fraction of the largest connected components S versus the fraction of removed nodes f is frequently used to investigate the robustness of a network [4], [54]. In Fig. 7, we draw the curves for both the real-world and artificial networks as constructed in Section II.

From Fig. 7, we can conclude that the PPI networks are very robust to random mutations, while they are very fragile to targeted mutations. With the increasing of fraction of randomly removed nodes, the curves S versus f decrease slowly, while with the increasing of fraction of highly connected nodes or nodes with high FPC values, the sizes of largest connected components decrease quickly. An interesting finding is that for the real-world networks, by comparing the results between targeted mutations by FPC and degree, we find that there are certain robustness under targeted mutations on nodes with the largest FPC. Furthermore, targeted mutations on less than 2% nodes with the largest FPC values can produce roughly similarly results as random mutations, which indicates

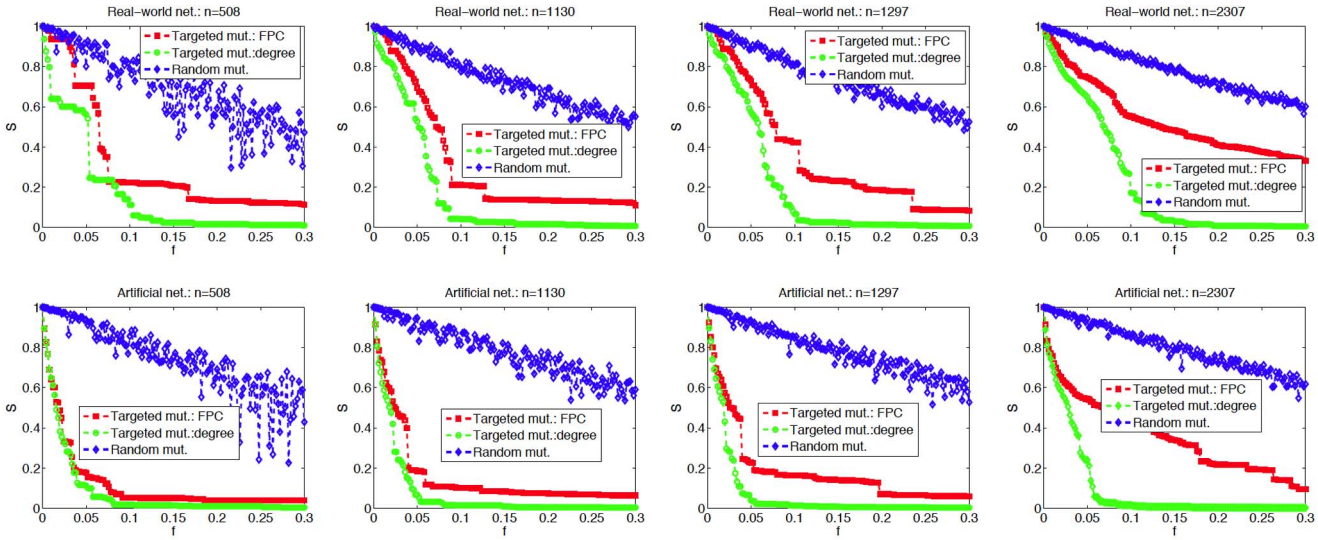

Fig. 7. Effect of targeted and random mutations on the PPI networks. The fractions of the largest connected component S versus the fraction of removed nodes f are plotted. Targeted mutations on proteins with large degree and with large FPC values have been considered.

that the structure of the PPI networks can not be severely destroyed by removing a few SDN. In real-world disease networks, it is crucial to keep network structures while curing disease, therefore, SDN identified by our method can be taken as potential drug targets. It is also observed that with the growth of network size, the curves S versus f obtained by the FPC decreases slower and slower, which indicates that targeted mutations on nodes with large FPC values can not severely destroy the network as mutations on nodes with large degree, it also indicates network size may have certain effect on mutation robustness.

V. Conclusion

Identifying important nodes in complex networks has been a popular topics in recent years. In this paper, based on some existing measures of node importance and the PCA method, we have proposed a new integrative node importance measure and identified SDN in some evolving PPI networks. The evolving networks origin from the same small networks, but with increased network size and evolved topological structures. The investigated real-world yeast PPI networks are based on existing databases and publications, while the artificial networks are constructed by the DD model. Under properly chosen parameter, the generated artificial networks can well mimic the statistical characteristics of the real-world PPI ones, such as they have similar average node degree, average path length, clustering coefficients and power-law exponents. Most interestingly, both real-world and artificial PPI networks consist of similar network motifs as building blocks.

By the proposed new indicator, we have successfully identified the SDP in real-world yeast and artificial PPI networks. Clustering analysis and rich-club analysis indicate that PPI networks have clear hierarchical structures, with only a few SDP, and the few SDP are highly connected. The proposed integrative measure FPC is closely correlated with the eigenvector centrality, the semi-local centrality, the network motif centrality, the degree and betweenness measures, which reveals the relations between SDN and important nodes identified by the other methods.

PPI networks evolve with time, which leads to the evolution of node importance. By performing numerical simulations on extensive artificial networks, we have also investigated the evolution characteristics of node importance. It is found that, after a long time evolution, most of the top-ranked nodes can maintain dominance, while a few nodes may become less and less dominant. Moreover, the less dominant a node is, the more rapidly it evolves to be much less dominant. Furthermore, it is also found that PPI networks are robust to random mutations while fragile to targeted mutations. Targeted mutations on SDN and high degree nodes can bring about similar yet different consequences, the networks can keep certain robustness under the first mutation strategy.

Due to space limitation, we have only considered the representative yeast PPI networks. It is noted that the Yeast PPI networks are the most frequently investigated bio-molecular networks, available data for the Yeast PPIs are with very high quality [23]–[31]. We note that the related methods and investigations can be easily extended to other types of networks, such as social networks and directed gene networks, metabolic networks, electrical networks. They can also be easily extended to the other organisms, such as the PPI networks for Human, E. coli and Xenopus laevis. Our investigations on evolving PPI networks can shed some light on the future applications of the evolving characteristics of complex networks, such as reengineering of particular bio-molecular networks for technological, synthetical or pharmacological purposes [56]–[59]. For example, in real-world disease-related gene networks or PPI networks, one may detect SDN and consider to take them as control targets during the treatment of disease, since the existing investigations have suggested that disease-related genes are characterized with certain structural features in the PPI networks [18]–[20], SDN are more appropriate to be chosen as potential control targets [60], targeted mutations on these nodes will not severely destroy the networks.

REFERENCES

[1] L. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[2] L. Freeman, "Centrality in social networks: Conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, 1979.

[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[4] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[5] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Netw.*, vol. 27, pp. 39–54, 2005.

[6] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir, "A model of Internet topology using k-shell decomposition," *Proc. Nat. Acad. Sci. USA*, vol. 104, pp. 11150–11154, 2007.

[7] P. Bonacich and P. Lloyd, "Eigenvector-like measures of centrality for asymmetric relations," *Social Netw.*, vol. 23, no. 3, pp. 191–201, 2001.

[8] C. Canali and R. Lancellotti, "A quantitative methodology based on component analysis to identify key users in social networks," *Int. J. Social Netw. Mining*, vol. 1, no. 1, pp. 27–50, 2012.

[9] D. Chen, L. Lü, M. Shang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica A*, vol. 391, pp. 1777–1787, 2012.

[10] S. Brin and L. Page, "The anatomy of a largescale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, pp. 107–117, 1998.

[11] L. Lü, Y. Zhang, C. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PLoS One*, vol. 6, no. art. e21202, 2011.

[12] F. Probst, D. K. L. Grosswiele, and D. K. R. Pfleger, "Who will lead and who will follow: Identifying influential users in online social networks," *Bus. Inform. Syst. Eng.*, vol. 5, pp. 179–193, 2013.

[13] O. Sporns, C. J. Honey, and R. Kötter, "Identification and classification of hubs in brain networks," *PLoS One*, vol. 2, no. 10, art. e1049, 2007.

[14] P. Bertolazzi, M. E. Bock, and C. Guerra, "On the functional and structural characterization of hubs in protein-protein interaction networks," *Biotechnol. Adv.*, vol. 31, no. 2, pp. 274–286, 2013.

[15] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nat. Phys.*, vol. 6, pp. 888–893, 2010.

[16] P. Wang, C. Tian, and J. Lu, "Identifying influential spreaders in artificial complex networks," *J. Syst. Sci. Complex.*, 2013, to be published.

[17] P. Wang, J. Lü, and X. Yu, "Identification of important nodes in directed biological networks: A method based on network motifs," *PLoS One*, 2014, revised version, currently under review, submitted for publication.

[18] E. Wang, A. Lenferink, and M. O'Connor-McCourt, "Cancer systems biology: Exploring cancer-associated genes on cellular networks," *Cell Mol. Life Sci.*, vol. 64, pp. 1752–1762, 2007.

[19] G. Östlund, M. Lindskog, and E. L. Sonnhammer, "Network-based identification of novel cancer genes," *Mol. Cell Proteom.*, vol. 9, pp. 648–655, 2010.

[20] X. Wang, N. Gulbahce, and H. Yu, "Network-based methods for human disease gene prediction," *Brief Funct. Genom.*, vol. 10, pp. 280–293, 2011.

[21] A. Wagner, "The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes," *Mol. Biol. Evol.*, vol. 18, no. 7, pp. 1283–1292, 2001.

[22] S. Wuchty, "Interaction and domain networks of yeast," *Proteom.*, vol. 2, no. 12, pp. 1715–1723, 2002.

[23] B. Schwikowski, P. Uetz, and S. Fields, "A network of protein-protein interactions in yeast," *Nat. Biotechnol.*, vol. 18, no. 12, pp. 1257–1261, 2000.

[24] H. Yu *et al.*, "High-quality binary protein interaction map of the Yeast interactome network," *Science*, vol. 322, no. 5898, pp. 104–110, 2008.

[25] Y. Jin, D. Turaev, T. Weinmaier, T. Rattei, and H. A. Makse, "The evolutionary dynamics of protein-protein interaction networks inferred from the reconstruction of ancient networks," *PLoS One*, vol. 8, no. 3, art. e58134, 2013.

[26] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "Biogrid: A general repository for interaction datasets," *Nucleic Acids Res.*, vol. 34, pp. D535–D539, 2006.

[27] W. E. Payne and J. I. Garrels, "Yeast Protein database (YPD): A database for the complete proteome of Saccharomyces cerevisiae," *Nucleic Acids Res.*, vol. 25, no. 1, pp. 57–62, 1997.

[28] H. W. Mewes *et al.*, "MIPS: Analysis and annotation of proteins from whole genomes in 2005," *Nucleic Acids Res.*, vol. 34, pp. D169–D172, 2006.

[29] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, and D. Eisenberg, "DIP: The database of interacting proteins," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 289–291, 2000.

[30] P. Uetz *et al.*, "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae," *Nature*, vol. 403, no. 6770, pp. 623–627, 2000.

[31] T. Ito *et al.*, "Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins," *Proc. Nat. Acad. Sci. USA*, vol. 97, pp. 1143–1147, 2000.

[32] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler, "A model of large-scale proteome evolution," *Adv. Complex Syst.*, vol. 5, pp. 43–54, 2002.

[33] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, "Modeling of protein interaction networks," *Complexus*, vol. 1, pp. 38–44, 2003.

[34] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, "Duplication-divergence model of protein interaction network," *Phys. Rev. E*, vol. 71, no. art. 061911, 2005.

[35] X. Wan, S. Cai, J. Zhou, and Z. Liu, "Emergence of modularity and disassortativity in protein-protein interaction networks," *Chaos*, vol. 20, no. art. 045113, 2010.

[36] C. Xu, Z. Liu, and R. Wang, "How divergence mechanisms influence disassortative mixing property in biology," *Physica A*, vol. 389, no. 3, pp. 643–650, 2010.

[37] D. Zhao, Z. Liu, and J. Wang, "Duplication: A mechanism producing disassortative mixing networks in biology," *Chin. Phys. Lett.*, vol. 24, pp. 2766–2768, 2007.

[38] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Adv. Phys.*, vol. 51, pp. 1079–1187, 2002.

[39] P. Wang, X. Yu, and J. Lü, "Identification of important nodes in artificial bio-molecular networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2014.

[40] R. Pastor-Satorras, E. Smith, and R. V. Solé, "Evolving protein interaction networks through gene duplication," *J. Theor. Biol.*, vol. 222, no. 2, pp. 199–210, 2003.

[41] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, pp. 824–827, 2002.

[42] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, "Network motifs in the transcriptional regulation network of Escherichia coli," *Nat. Genet.*, vol. 31, pp. 64–68, 2002.

[43] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosoph. Mag.*, vol. 2, pp. 559–572, 1901.

[44] W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*. Berlin, Germany: Springer-Verlag, 2012.

[45] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London, U.K.: Chapman & Hall/CRC, 2007.

[46] P. Wang, J. Lü, and M. Ogorzalek, "Global relative parameter sensitivities of the feed-forward loops in genetic networks," *Neurocomput.*, vol. 78, pp. 55–165, 2012.

[47] P. Wang, J. Lü, Y. Zhang, and M. Ogorzalek, "Global relative input-output sensitivities of the feed-forward loops in genetic networks," in *Proc. 31th Chin. Control Conf.*, Jul. 25–27, 2012, pp. 7376–7381.

[48] P. Wang and J. Lü, "Control of genetic regulatory networks: Opportunities and challenges," *Acta. Automat. Sinica*, vol. 39, no. 12, pp. 1969–1979, 2013.

[49] P. Wang, J. Lü, Y. Zhang, and M. Ogorzalek, "Intrinsic noise induced state transition in coupled positive and negative feedback genetic circuit," in *Proc. IEEE Int. Conf. Syst. Biol.*, Sep. 2–4, 2011, pp. 356–361.

[50] M. Ashburner *et al.*, "Gene ontology: Tool for the unification of biology. The Gene Otology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, 2000.

[51] D. E. Sterner *et al.*, "Functional organization of the yeast SAGA complex: Distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction," *Mol. Cell Biol.*, vol. 19, no. 1, pp. 86–98, 1999.

[52] E. Koutelou, C. L. Hirsch, and S. Y. R. Dent, "Multiple faces of the SAGA complex," *Curr. Opin. Cell Biol.*, vol. 22, pp. 374–382, 2010.

[53] P. Jancura and E. Marchiori, "A survey on evolutionary analysis in PPI networks," InTech, Protein interaction/book 2, 2011.
[54] R. Albert, H. Jeong, and A. L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–82, 2000.
[55] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, 2001.
[56] B. Chen, W. Wu, Y. Wang, and W. Li, "On the robust circuit design schemes of biochemical networks: Steady-state approach," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 91–104, 2007.
[57] B. Chen and P. Chen, "Robust engineered circuit design principles for stochastic biochemical networks with parameter uncertainties and disturbances," *IEEE Trans. Biomed. Circuits Syst.*, vol. 2, no. 2, pp. 114–132, 2008.
[58] M. Gu and S. Chakrabartty, "FAST: A framework for simulation and analysis of large-scale protein-silicon biosensor circuits," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 4, pp. 451–459, Aug. 2013.
[59] F. Wu, "Global and robust stability analysis of genetic regulatory networks with time-varying delays and parameter uncertainties," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 391–398, Aug. 2011.
[60] S. Roy, "Systems biology beyond degree, hubs and scale-free networks: The case for multiple metrics in complex networks," *Syst. Synth. Biol.*, vol. 6, pp. 31–34, 2012.

Pei Wang received the M.Sc. and Ph.D. degrees in computational mathematics from the School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, China, in 2009 and 2012, respectively.

Currently, he is a Lecturer at the School of Mathematics and Information Sciences, Henan University, Kaifeng, China. During August 2013 (and upcoming in August 2014), he was a Visiting Research Fellow at the School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology (RMIT), Melbourne, Victoria, Australia. His research interests include systems biology, complex systems and networks, and chaotic dynamical systems.

Dr. Wang is a reviewer of several international journals, including the *International Journal of Bifurcation and Chaos*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, *Neural Computing & Applications*, *Acta Automatica Sinica*, the *Journal of Systems Science and Complexity*. He is an editorial board member of the *Chinese Journal of Mathematics*.

Xinghuo Yu (F'08) received the B.Eng. and M.Eng. degrees from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree from Southeast University, Nanjing, China, in 1982, 1984, and 1988, respectively.

Currently, he is with the Royal Melbourne Institute of Technology (RMIT), Melbourne, Vicotoria, Australia, where he is the Founding Director of the RMIT Platform Technologies Research Institute. His research interests include variable structure and nonlinear control, complex and intelligent systems, and applications.

Dr. Yu is an IEEE Distinguished Lecturer and Vice President (Publications) of the IEEE Industrial Electronics Society. He has received a number of awards and honors for his contributions, including the 2013 Dr.-Ing. Eugene Mittlemann Achievement Award of the IEEE Industrial Electronics Society and 2012 IEEE Industrial Electronics Magazine Best Paper Award. He is a Fellow of the Institution of Engineering and Technology (U.K.), the Institution of Engineers, Australia, and the Australian Computer Society.

Jinhu Lü (F'13) received the Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, China, in 2002.

Currently, he is a Professor in the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He was a Visiting Fellow at Princeton University, Princeton, NJ, USA, from 2005 to 2006. He is the author of three research monographs and more than 110 SCI journal papers published in the fields of complex networks and complex systems, nonlinear circuits and systems, receiving more than 6000 SCI citations with h-index 40. He also has two granted patents.

Dr. Lü serves as a member of the Fellows Evaluating Committee of the IEEE Circuits and Systems Society. He served and is serving as Editor in various ranks for 11 SCI journals including IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: EXPRESS BRIEFS, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *International Journal of Bifurcation and Chaos*, and *Asian Journal of Control*. He received the prestigious State Natural Science Award twice from the Chinese government in 2008 and 2012, respectively, the 9th Guanghua Engineering Science and Technology Award from the Chinese Academy of Engineering in 2012, the 11th Science and Technology Award for Youth of China and the Australian Research Council Future Fellowships Award in 2009. Moreover, he was awarded a National Natural Science Fund for Distinguished Young Scholars and selected by the 100 Talents Program of the Chinese Academy of Sciences.