Spectrum Analysis Based Method for Dynamics and Collective Analysis of Protein-Protein Interaction Networks

Yi-Zhen Shen[1], Yong-Sheng Ding[1,2*], Quan Gu[1]
1) College of Information Sciences and Technology
2) Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education
Donghua University, Shanghai 201620, P. R. China
*Corresponding to: ysding@dhu.edu.cn (Y.S. Ding).

Abstract. The importance of understanding biological interaction networks has fueled the development of numerous interaction data generation techniques, databases and prediction tools. Generation of high-confident interaction networks formulates the first step towards the study for protein–protein interactions (PPI). A number of experimental methods, based on distinct, physical principles have been developed to identify PPI such as the yeast two-hybrid method (Y2H). In this work, we focus on one example of biological networks, namely the yeast protein interaction network (YPIN). In YPIN, we design and implement a computational model that captures the discrete and stochastic nature of protein interactions. In this model, we apply spectrum analysis method to the variance of the protein nodes which play an important role in the PPI networks, which can show the topology structure of dynamic and collective performances of PPI networks. We take YPIN, such as 48 "quasi-cliques" and 6 "quasi-bipartites" separated from 11855 yeast PPI networks with 2617 proteins, as an example and apply spectrum analysis to show the topology structure of dynamic and collective analysis of PPI networks and the performances. The obtained results may be valuable for deciphering unknown protein functions, determining protein complexes, and inventing drugs.

Key words: protein–protein interactions, evidence networks, visual analytics, exploratory analysis, spectrum analysis

1 Introduction

Protein–protein interaction (PPI) networks are complex entities generally containing thousands of interactions. Prediction tools as well as experimental techniques aim at assigning quantitative metrics to the interaction edges within such networks. In most cases, the evidence about these interactions can be incomplete and associated with uncertainty. As a result, human judgment and expertise has to be exercised while deriving a set of high-confidence interactions after assessing each source of data. Most researchers use some sort of a voting mechanism to filter interactions, or use predictions from a method that has the most confidence in. This process is not only tedious but also subjective in nature. Methods are needed to examine this multi-dimensional, multi-source data in an automated and timely manner. The lack of computational tools facilitating such analysis of evidence from multiple prediction and experimental sources such as Gene Neighbourhood-GN, Gene Cluster-GC, Phylogenetic Profiles-PP, BIND [1], and DIP [2] is the motivation behind this study.

The complex network of relationships involved in biological systems such as PPI, small-molecule metabolism, and gene regulation, can now be studied in increasing detail [3]. There have been a considerable number of studies into general biological network properties, such as their scale-free nature.

M. Chetty, S. Ahmad, A. Ngom, S. W. Teng (Eds.): PRIB2008,
Supplementary Conf. Proc., pp.147-158, 2008.
(c) PRIB 2008

The representation of protein structures as small-world networks has recently become an interesting approach to study a variety of problems associated to protein function and structure. The instances are the identification of key residues involved in the protein folding mechanism [4] and the correlation between the topological properties of protein conformations. Besides these, they also include the kinetic ability to fold [5] and [6], or the identification of functional sites in protein structures [7] among other examples.

An interesting application of small-world networks would be the representation of protein–protein complexes as such networks, in order to elucidate different structural characteristics associated with the presence of residues that contribute the most to the binding free energy (hot spots), which are unevenly distributed at the binding interface [8]. Although different approaches involving sequence and structural information or energetic calculations have been proposed to study and predict hot spots of binding free energy [9], the small-world representation of protein–protein complexes could give another complementary view on this problem.

In the past three years, several researches on high flux interaction have been developed, such as yeast bi-hybrid system [10], protein purification method based on mass spectrum analysis [11] and [12], expression spectrum analysis method with relative information [13] and [14], genetic interaction networks method [15], and other interaction prediction methods based on gene relativity calculation model, gene amalgamation and abruption [16], gene neighbor [17] and gene with together appearance [18] and [19] etc. In several life-forms (such as S. cerevisiae, Caenorhabditis elegans and Helicobacter pylori), massive interaction data were produced. These high flux data open a door for knowing more generally biological phenomena of cells. Afterwards, several methods were successfully applied in this field study. For instance, Schwikowski (2000) and Hishigaki (2001) made use of interaction neighbor to predict the unknown protein function. Ge and Liu (2001) proved that the proteins with similar expression spectrum interacted with each other for the first time. Fraser and Hirsh (2002) opened out that the proteins with conservative interaction had negative correlation with their breaking rate. All of these studies indicate that S. cerevisiae interaction networks maybe have different characters with other complex networks.

Interaction topology mode is an important start to study biological function information of protein, so we need to develop some methods to analyze PPI networks. In this paper, we propose spectrum analysis method, which has been successfully applied in other fields [20], to study PPI networks, identify topology mode (quasi-cliques and quasi-bipartites) of PPI networks, and analyze dynamics and collective performances of PPI networks.

The paper is organized as follows: Section 2 discusses the computational model that captures the discrete and stochastic nature of protein interactions. Section 3 applies spectrum analysis to show the topology structure of dynamic and collective performances of PPI networks and analyzes the variance of the protein node that plays an important role in the PPI network. Section 4 concludes the paper by discussing spectrum analysis method for dynamics and collective performances of PPI networks.

2 Methods

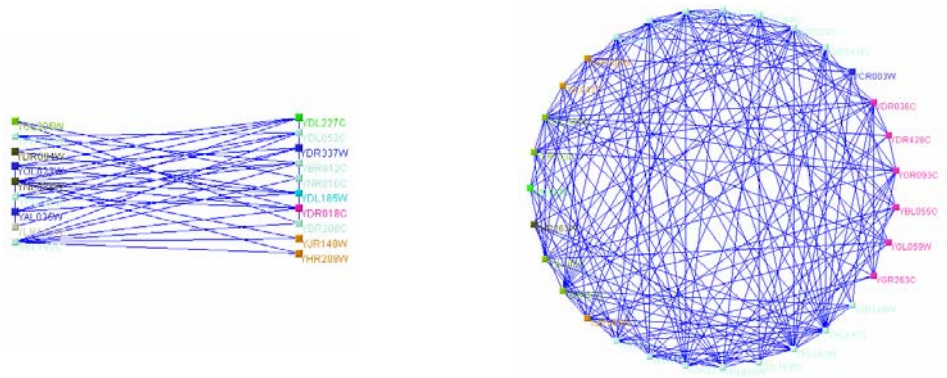
2.1 Yeast protein interaction network

In this blooming field, systems-wide analyses of PPI has taken center stage. A number of experimental methods, based on distinct, physical principles have been developed to identify PPI such as the yeast two-hybrid method (Y2H), affinity purification-mass spectrometry (AP-MS) approaches and protein

microarrays. The primary modus operandi to identify and map binary, physical PPI is the Y2H assay.

The Y2H system is reliably and easily adapted to high-throughput technology and provides the most widely used method for studies of PPI. This method is easy to implement, amenable to automation and relatively cheap. Nearly any protein, including most therapeutically relevant target proteins, expressed as Bait protein fused to a DNA-binding domain can be screened against the full complement of a proteome, expressed as individual Prey proteins fused to a transcriptional activator domain. The Y2H method has been used to generate global PPI networks for the bacterial strain Helicobacter pylori, the yeast Saccharomyces cerevisiae, the fruit fly Drosophila melanogaster, the worm Caenorhabditis elegans and recently a significant fraction of the human proteome.

We take yeast protein interaction network (YPIN), such as “quasi-cliques” and “quasi-bipartites” separated from yeast PPI networks, as an example and apply spectrum analysis to show the topology structure of dynamic and collective analysis of PPI networks. We find that these connotative structures correspond to the biological functional modules. Making use of spectrum analysis, 48 “quasi-cliques” and 6 “quasi-bipartites” are separated from the 11855 yeast PPI networks with 2617 proteins. The typical “quasi-cliques” and “quasi-bipartites” are as shown in Fig. 1. We choose a typical yeast PPI network from the 48 “quasi-cliques” as an example to analyze the dynamics and collective performances.



(a) quasi-bipartite (b) quasi-clique

Fig. 1. Topology structure of the typical yeast PPI network.

2.2 Spectrum Analysis

Spectrum analysis is an efficient method which can analyze the deep-seated structure of a great deal of data relationship. As a famous example, Gibson, Kleinberg and Raghavan acquired achievement in the field of www (World Wide Web) linking structure and information mining [21] and [22]. As we all known, www is linked by more and more web pages. Besides www structure complexity, spectrum analysis method was also applied to find out some data sources, such as “authoritative” and “hub”.

As to our computational model, spectrum analysis is applied to identify the interesting topology structure of dynamic and collective analysis of PPI networks. Through this method, PPI network is expressed as a non-direction plot $G(N,E)$, where $N=\{P_1,P_2,...,P_n\}$ is a node set including all protein nodes, $E=\{(P_i,P_j)\,|\,\text{the interaction between } P_i \text{ and } P_j\}$ is defined as a line set. A symmetrical $n\times n$ linking matrix can be defined as $A=(a_{ij})$, where

$$a_{ij} = \begin{cases} 1 & (P_i, P_j) \in E \\ 0 & (P_i, P_j) \notin E \end{cases} \quad \textbf{(1)}$$

From the spectrum essence analysis, the linking matrix A is an important measurement according to the node attributes in PPI networks. Scoring every protein node can be considered to show the node "importance", such as X. A high scored protein node can add the score of the nodes through interaction, which means that the two interaction protein nodes can add their scores. We apply a cycle to define their scores

$$\Delta X_i = \sum_{j=1}^{n} a_{ij} \times X_j \quad \textbf{(2)}$$

Gibson and others [23] made use of iterative algorithm to introduce a method to break of this cycle. What is interesting is that X_i is always a special protein node no matter what the initial value is. Moreover, because of the symmetrical matrix A and the orthogonal eigenvector, the corresponding attributes may be orthogonal. In other words, each eigenvector may indicate a special attribute which does not belong to other eigenvectors.

We apply this spectrum analysis algorithm to construct a typical "quasi-cliques" PPI network shown in Fig. 2. This PPI network shows its topology structure, which can be used to analyse the dynamic and collective performances of PPI networks.

2.3 Small-world Networks

Up to now, networks of coupled dynamical systems have been used to model biological oscillators, Josephson junction arrays, excitable media, neural networks, spatial games, genetic control networks and many other self-organizing dynamical systems [24]. Generally, the topology connection is proposed to be either completely regular or completely random. However, many biological, technological and social networks lie somewhere between the two extremes, which is called small-world networks. Models of dynamical systems, especially dynamic and collective PPI networks, are small-world coupling, which display enhanced signal-propagation speed, computational power, and synchronizability. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.

To interpolate between regular and random networks, we propose a random rewiring procedure. Starting from a ring lattice with n nodes and k edges per node, we rewire each edge at random with probability p-value, which indicates the proportion of the real connection number and the total connection number. This construction allows us to "tune" the graph between regularity (p=0) and disorder (p=1), and thereby to probe the intermediate region 0<p<1, about which little is known. We propose that the whole PPI network is composed with "quasi-cliques", "quasi-bipartites" and other proteins we have not predicted.

P-value indicates the topology structure of PPI networks, which is useful for us to analyse the dynamic and collective performances of PPI networks. Because of the p-value definition, we can propose that the protein nodes with high p-value have large effect on PPI networks. On contrary, the protein nodes with low p-value have little effect on PPI networks.

3 Results and Discussions

3.1 The Original PPI Network

We design and implement a computational model that captures the discrete and stochastic nature of protein interactions. In our model, spectrum analysis is applied to show the topology structure of PPI networks, such as “quasi-cliques” and “quasi-bipartites”. We find that these connotative structures correspond to the biological functional modules. Making use of spectrum analysis, 48 “quasi-cliques” and 6 “quasi-bipartites” are separated from the 11855 yeast PPI networks with 2617 proteins.

A typical PPI network is chosen from the 48 “quasi-cliques” shown in Fig. 2 as an example to analyze the dynamics and collective performances. Based on the analysis of the PPI, we find that this network is neither a regular network nor a random network. As a result, the PPI network is a small-world network [25].

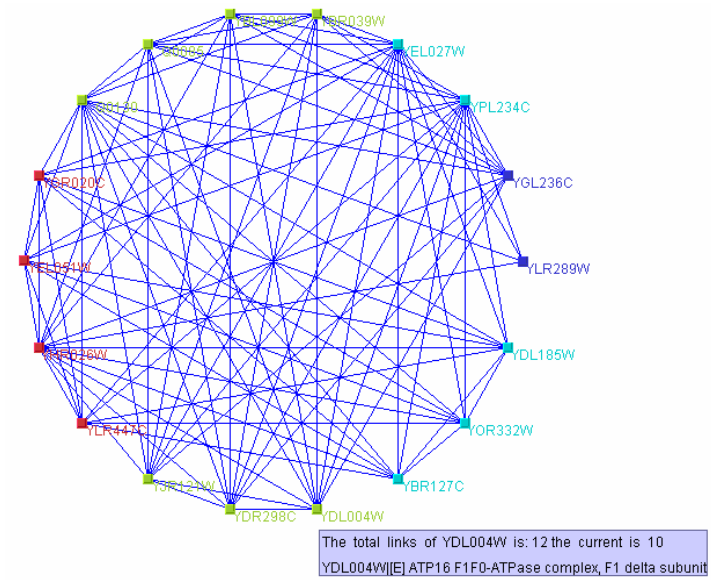


Fig. 2. A typical “quasi-cliques” PPI network.

In this small-world PPI network, we analyze the 18 protein nodes and the whole network. The total links is the number of the total edges connecting one protein node and other ones in the 11855 yeast PPI networks with 2617 proteins. The current links is the number of the edges connecting one protein node and other ones in current small-world protein network. Degree of the protein node is the same meaning as the current links. Degree of the PPI network is the average of the total protein nodes degrees, which displays the collective extent of the PPI network. p-value here is the same meaning as the one in small-world network, as defined above.

The parameters of the PPI network are listed in Table 1. The bolded proteins, YHR026W, Q0130, YEL027W and YPL234C, have the highest p-value. The bolded protein, YLR289W has the lowest p-value, which will be used to analyze the dynamics and collective performances of the PPI network.

Table 1. Parameters of the PPI network.

Protein	Total links	Current links	Degree	p-value ($0 \le p \le 1$)	Protein attributes
YEL051W	12	9	9	0. 5294	VMA8 H+-ATP synthase V1 domain 32 KD subunit, vacuolar
YHR026W	16	14	14	**0.8235**	PPA1 H+-ATPase 23 KD subunit, vacuolar
YGR020C	12	9	9	0.5294	VMA7 H+-ATPase V1 domain 14 kDa subunit, vacuolar
YLR447C	18	9	9	0.5294	VMA6 H+-ATPase V0 domain 36 KD subunit, vacuolar
Q0130	15	14	14	**0.8235**	OLI1 F1F0-ATPase complex, F0 subunit 9
YJR121W	12	10	10	0.5882	ATP2 F1F0-ATPase complex, F1 beta subunit
Q0085	10	10	10	0.5882	ATP6 F1F0-ATPase complex, FO A subunit
YDR298C	12	10	10	0.5882	ATP5 F1F0-ATPase complex, OSCP subunit
YBL099W	12	10	10	0.5882	ATP1 F1F0-ATPase complex, F1 alpha subunit
YDL004W	12	10	10	0.5882	ATP16 F1F0-ATPase complex, F1 delta subunit
YBR039W	12	10	10	0.5882	ATP3 F1F0-ATPase complex, F1 gamma subunit
YBR127C	13	9	9	0.5294	VMA2 H+-ATPase V1 domain 60 KD subunit, vacuolar
YEL027W	15	14	14	**0.8235**	CUP5 H+-ATPase V0 domain 17 KD subunit, vacuolar
YOR332W	13	9	9	0.5294	VMA4 H+-ATPase V1 domain 27 KD subunit, vacuolar
YPL234C	16	14	14	**0.8235**	TFP3 H+-ATPase V0 domain 17 KD subunit, vacuolar
YDL185W	36	9	9	0.5294	TFP1 H+-ATPase V1 domain 69 KD catalytic subunit, vacuolar
YGL236C	13	10	10	0.5882	MTO1 strong similarity to gidA E. coli protein
YLR289W	12	4	4	**0.2353**	GUF1 strong similarity to E. coli elongation factor-type GTP-binding protein lepa

3.2 Taken out the protein nodes with p-value maximum and minimum

In Table 1, we aim at the protein nodes with p-value maximum and p-value minimum to analyze the dynamics and collective performances of the PPI network. The protein node YHR026W with p-value maximum is taken out from the network. The degree of the network changes from 10 to 9. The topology structure of the PPI network taken out the protein node YHR026W with p-value maximum also changes. The changed protein parameters of the PPI network are listed in Table 2.

We compare the protein networks before taken out the protein node YHR026W with p-value maximum from the network and the one after variance. In Fig. 3, except the protein node YHR026W itself, we can see that there are 14 of 17 protein nodes' p-value changed with different degrees.

In the same way, the protein node YLR289W with p-value minimum is taken out from the network to analyze the dynamics and collective performances of the PPI network. The degree of the network does not change. The topology structure of the PPI network taken out the protein node YLR289W with p-value minimum also changes. The changed protein parameters of the PPI network are listed in Table 3.

Table 2. Parameters of the PPI network after variance.

Protein	Total links	Current links	Degree	p-value ($0 \le p \le 1$)	Protein attributes
YEL051W	12	8	8	0.4706	VMA8 H+-ATP synthase V1 domain 32 KD subunit, vacuolar
YGR020C	12	8	8	0.4706	VMA7 H+-ATPase V1 domain 14 kDa subunit, vacuolar
YLR447C	18	8	8	0.4706	VMA6 H+-ATPase V0 domain 36 KD subunit, vacuolar
YJR121W	12	9	9	0.5294	ATP2 F1F0-ATPase complex, F1 beta subunit
Q0085	10	9	9	0.5294	ATP6 F1F0-ATPase complex, FO A subunit
YDR298C	12	9	9	0.5294	ATP5 F1F0-ATPase complex, OSCP subunit
YBL099W	12	9	9	0.5294	ATP1 F1F0-ATPase complex, F1 alpha subunit
YDL004W	12	9	9	0.5294	ATP16 F1F0-ATPase complex, F1 delta subunit
YBR039W	12	9	9	0.5294	ATP3 F1F0-ATPase complex, F1 gamma subunit
YBR127C	13	8	8	0.4706	VMA2 H+-ATPase V1 domain 60 KD subunit, vacuolar
YOR332W	13	8	8	0.4706	VMA4 H+-ATPase V1 domain 27 KD subunit, vacuolar
YDL185W	36	8	8	0.4706	TFP1 H+-ATPase V1 domain 69 KD catalytic subunit, vacuolar
YGL236C	13	9	9	0.5294	MTO1 strong similarity to gidA E. coli protein
YLR289W	12	3	3	0.1765	GUF1 strong similarity to E. coli elongation factor-type GTP-binding protein lepa

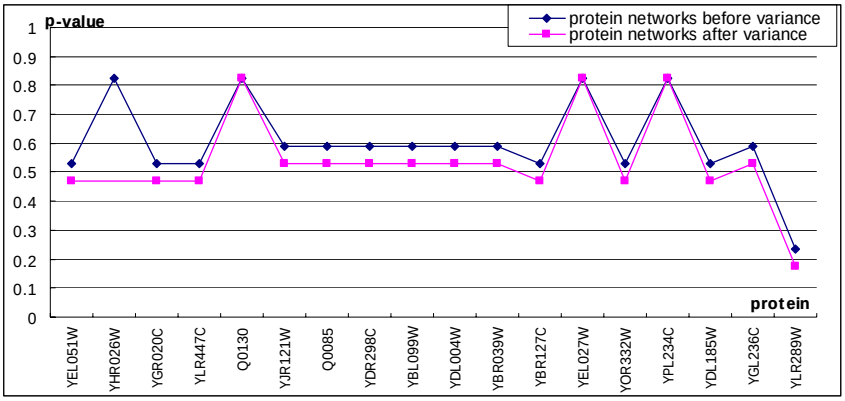


Fig. 3. Comparison between the original PPI network and the one taken out the protein node YHR026W with p-value maximum.

We compare the protein networks before taken out the protein node YLR289W with p-value minimum from the network and the one after variance. In Fig. 4, except the protein node YLR289W itself, we can see that there are only 4 of 17 protein nodes' p-value changed with different degrees.

Table 3. Parameters of the PPI network after variance.

Protein	Total links	Current links	Degree	p-value ($0 \le p \le 1$)	Protein attributes
YHR026W	16	13	13	0.7647	PPA1 H+-ATPase 23 KD subunit, vacuolar
Q0130	15	13	13	0.7647	OLI1 F1F0-ATPase complex, F0 subunit 9
YEL027W	15	13	13	0.7647	CUP5 H+-ATPase V0 domain 17 KD subunit, vacuolar
YPL234C	16	13	13	0.7647	TFP3 H+-ATPase V0 domain 17 KD subunit, vacuolar

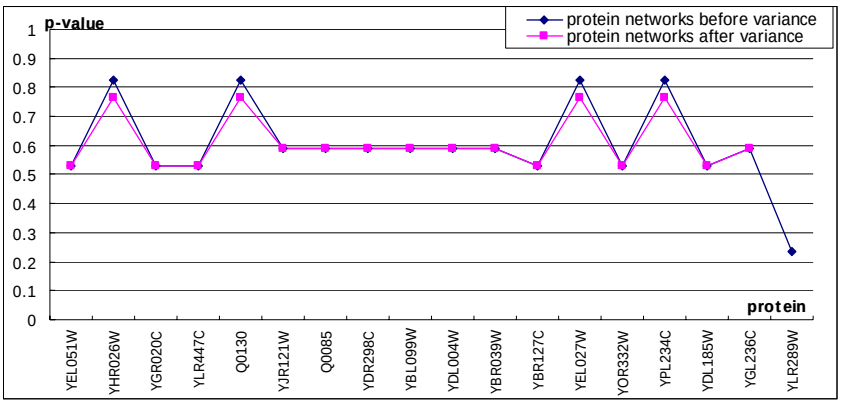


Fig. 4. Comparison between the original PPI network and the one taken out the protein node YLR289W with p-value minimum.

The percent of changed proteins of the PPI network taken out the protein node YHR026W with p-value maximum is 82.35%. However, the percent of changed proteins of the PPI network taken out the protein node YLR289W with p-value minimum is only 23.53%. We can conclude that the protein node YHR026W with p-value maximum taken out affects the dynamics and collective performances of the PPI network more than the protein node YLR289W with p-value minimum taken out.

3.3 Substituted the protein nodes with p-value maximum and minimum

Table 4. Parameters of the PPI network after variance.

Protein	Total links	Current links	Degree	p-value ($0 \le p \le 1$)	Protein attributes
YEL051W	12	8	8	0.4706	VMA8 H+-ATP synthase V1 domain 32 KD subunit, vacuolar
YHR020W	12	6	6	0.3529	strongly similarity to human glutamyl-prolyl-tRNA synthetase and fruit fly multifunctional aminoacyl-tRNA synthetase
YGR020C	12	8	8	0.4706	VMA7 H+-ATPase V1 domain 14 kDa subunit, vacuolar
YLR447C	18	8	8	0.4706	VMA6 H+-ATPase V0 domain 36 KD subunit, vacuolar
Q0130	15	15	15	0.8824	OLI1 F1F0-ATPase complex, F0 subunit 9
Q0085	10	9	9	0.5294	ATP6 F1F0-ATPase complex, FO A subunit
YDR298C	12	9	9	0.5294	ATP5 F1F0-ATPase complex, OSCP subunit
YBL099W	12	9	9	0.5294	ATP1 F1F0-ATPase complex, F1 alpha subunit
YDL004W	12	9	9	0.5294	ATP16 F1F0-ATPase complex, F1 delta subunit
YBR039W	12	9	9	0.5294	ATP3 F1F0-ATPase complex, F1 gamma subunit
YBR127C	13	8	8	0.4706	VMA2 H+-ATPase V1 domain 60 KD subunit, vacuolar
YEL027W	15	15	15	0.8824	CUP5 H+-ATPase V0 domain 17 KD subunit, vacuolar
YOR332W	13	8	8	0.4706	VMA4 H+-ATPase V1 domain 27 KD subunit, vacuolar
YPL234C	16	15	15	0.8824	TFP3 H+-ATPase V0 domain 17 KD subunit, vacuolar
YDL185W	36	8	8	0.4706	TFP1 H+-ATPase V1 domain 69 KD catalytic subunit
YLR289W	12	4	4	0.2353	GUF1 strong similarity to E. coli elongation factor-type GTP-binding protein lepa

The protein node YHR020W is substituted for YHR026W with p-value maximum to analyze the dynamics and collective performances of the PPI network. The degree of the network also changes from 10 to 9. The topology structure of the PPI network substituted the protein node YHR020W for the protein node YHR026W with p-value maximum also changes. The changed protein parameters of the PPI network are listed in Table 4.

We compare the protein networks before substituted the protein node YHR020W for YHR026W with p-value maximum and the one after variance. In Fig. 5, except the protein node YHR020W itself, we can see that there are 15 of 17 protein nodes' p-value changed with different degrees.

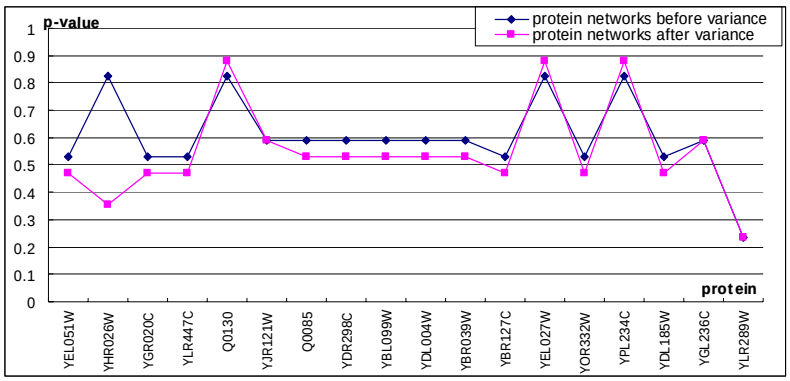


Fig. 5. Comparison between the original PPI network and the one substituted the protein node YHR020W for the protein node YHR026W with p-value maximum.

In the same way, the protein node YLR244C is substituted for YLR289W with p-value minimum to analyze the dynamics and collective performances of the PPI network. The degree of the network does not change. The topology structure of the PPI network substituted the protein node YLR244C for YLR289W with p-value minimum also changes. The changed protein parameters of the PPI network are listed in Table 5.

Table 5. Parameters of the PPI network after variance.

Protein	Total links	Current links	Degree	p-value ($0 \le p \le 1$)	Protein attributes
YHR026W	16	13	13	0.7647	PPA1 H+-ATPase 23 KD subunit, vacuolar
Q0130	15	13	13	0.7647	OLI1 F1F0-ATPase complex, F0 subunit 9
YEL027W	15	13	13	0.7647	CUP5 H+-ATPase V0 domain 17 KD subunit, vacuolar
YPL234C	16	13	13	0.7647	TFP3 H+-ATPase V0 domain 17 KD subunit, vacuolar
YLR244C	31	0	0	0	MAP1 methionine aminopeptidase, isoform 1

We compare the protein networks before substituted the protein node YLR244C for YLR289W with p-value minimum and the one after variance. In Fig. 6, except the protein node YLR244C itself, we can see that there are only 4 of 17 protein nodes' p-value changed with different degrees.

The percent of changed proteins of the protein-protein network substituted the protein node YHR020W for the protein node YHR026W with p-value maximum is 88.24%. However, the percent of changed proteins of the protein-protein network substituted the protein node YLR244C for YLR289W with p-value minimum is only 23.53%. We can conclude that the protein node YHR026W with p-value maximum substituted for affects the dynamics and collective performances of the PPI

network is more than the protein node YLR289W with p-value minimum substituted for.

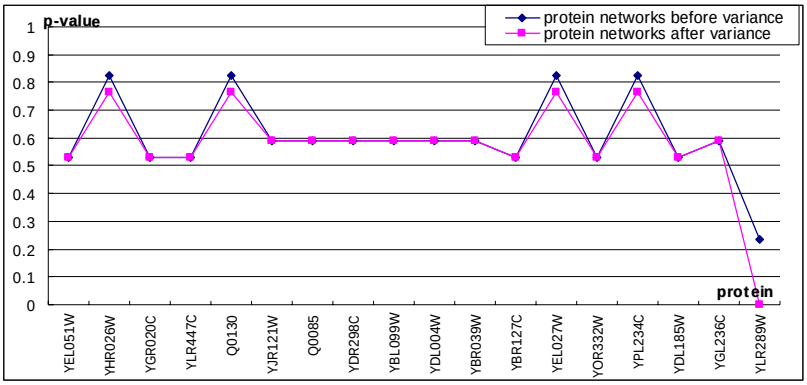


Fig. 6. Comparison between the original PPI network and the one substituted the protein node YLR244C for YLR289W with p-value minimum.

3.4 Effect on PPI Network

PPI networks for the yeast resulting from different high-throughput studies have been shown to have scale-free degree distributions, and thus these PPI networks are claimed to be "scale-free" [26].

We analyze the degree of PPI network for the above instances, and also compare the results in Table 6. Consequently, we could know the scale-free nature (power-law) of PPI network. We can find out that the protein node with p-value maximum acts an important role in the PPI network, because of the degree of PPI network alters from origin 10 to 9, which indicates the effect of its variance to the scale-free nature.

Table 6. Comparison of the degree of PPI network.

	Original	Taken out YHR026W	Taken out YLR289W	YHR026W → YHR020W	YLR289W→ YLR244C
Degree of PPI network	10	9	10	9	10

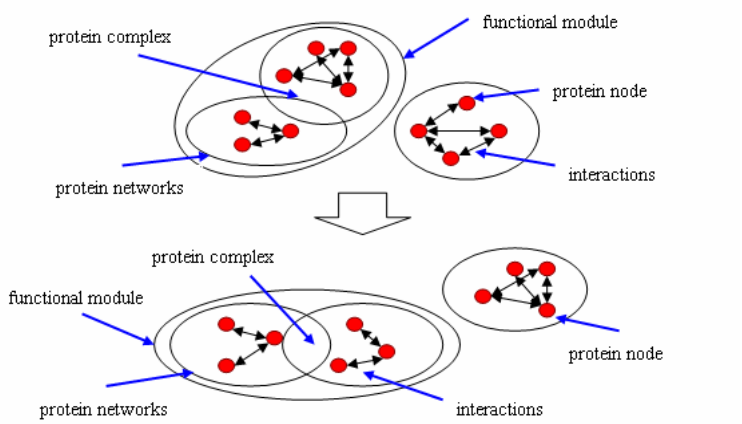


Fig. 7. The effect that the important protein nodes bring to the whole PPI network.

Spectrum analysis is applied in this paper to show the topology structure of dynamic and collective analysis of PPI networks, such as “quasi-cliques” and “quasi-bipartites”. We find that these connotative structures correspond to the biological functional modules. Making use of spectrum analysis, 48 “quasi-cliques” and 6 “quasi-bipartites” are separated from the 11855 yeast PPI networks with 2617 proteins.

Now, our work is to analyses the dynamic and collective variance of the PPI network, based on the topology structure. Some protein nodes in the biological network act an important role because their variance will alter the topology structure of the whole PPI network. The alteration process is showed in Fig. 7.

4 Conclusions

In this paper, we design and implement a computational model that captures the discrete and stochastic nature of protein interactions. In our model, we apply spectrum analysis to construct “quasi-cliques” PPI networks and show the topology structure of dynamic and collective performances of PPI networks. The YPIN has high exact rate in the note of protein nodes, which is about 87.5%. However, in random protein networks, the rate is only 2.1%. Therefore, we choose a typical PPI network from the 48 “quasi-cliques” in YPIN as an example to analyze the dynamics and collective performances of PPI networks.

We find that these connotative structures correspond to the biological functional modules, which has small-world network performances. In our work, we focus on one example of biological networks, namely the YPIN. In YPIN, we analyze the variance of the protein node that plays an important role in the PPI network. Generation of high-confident interaction networks formulates the first step towards further researches. So the results are valuable for deciphering unknown protein functions, determining protein complexes, and inventing drugs. Especially, in the field of inventing drugs, we could make use of medication to copy the important protein nodes if the disease is caused by the variance of the important protein nodes.

Acknowledgments

This work was supported in part by the Key Project of the National Nature Science Foundation of China (No. 60534020), the National Nature Science Foundation of China (No. 60775052, 70701009), Specialized Research Fund for the Doctoral Program of Higher Education from Ministry of Education of China (No. 20060255006), Shanghai Talent Developing Foundation, and Specialized Foundation for Excellent Talent from Shanghai.

Dr. Y.-S. Ding thanks to Prof. K.-C. Chou at Gordon Life Science Institute, San Diego, CA 92130, USA for his valuable suggestion and discussion. The authors wish to thank the anonymous reviewers, whose constructive comments are very helpful for improving the presentation of this study.

References

1. Bader, G. D. and Hogue, C. W., An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinformatics 4(1): 1471-2105, 2003.
2. Antonio, D. S., Hirotomo, F. and Paul O., Topology of small-world networks of protein–protein

complex structures, Bioinformatics, 21(8): 1311-1315, 2005.
3. Ren, L. H., Ding, Y. S., Shen, Y. Z. and Zhang, X. F., Multi-agent-based bio-network for systems biology: Protein-protein interaction network as an example, Amino Acid, 2008, in press.
4. Vazquez, A. and Flammini, A., Global protein function prediction from protein-protein interaction networks, Nat Biotechnol 21(6): 697-700, 2003.
5. Drysdale, R. A. and Crosby, M. A., FlyBase: genes and gene models, Nucleic Acids Res 33 Database Issue: D390-5, 2005.
6. Ghaemmaghami, S. and Huh, W. K., Global analysis of protein expression in yeast, Nature 425(6959): 737-41, 2003.
7. Ozier, O. and Amin, N., Global architecture of genetic interactions on the protein network, Nat Biotechnol 21(5): 490-491, 2003.
8. Ali, R. A., Pelin, A., and Canan, B., Small-world communication of residues and significance for protein dynamics, Biophysical Journal 86(1): 85-91, 2004.
9. Deng, M. and Tu, Z., Mapping Gene Ontology to proteins based on protein-protein interaction data, Bioinformatics 20(6): 895-902, 2004.
10. Bader, J. S. and Chaudhuri, A., Gaining confidence in high-throughput protein interaction networks, Nat Biotechnol 22(1): 78-85, 2004.
11. Diao, Y. B., Li, M. L., Fenga, Z. N., Yina, J. J. and Pana, Y., The community structure of human cellular signaling network, Journal of Theoretical Biology 247(4): 608-15, 2007.
12. Qin, H. and Lu, H. H., Evolution of the yeast protein interaction network, Proc Natl Acad Sci U S A 100(22): 12820-12824, 2003.
13. Bader, G. D. and Betel, D., BIND: the biomolecular interaction network database, Nucleic Acids Res 31(1): 248-50, 2003.
14. Pržulj, N., Wigle, D.A. and Jurisical, I., Functional topology in a network of protein interactions, BMC Bioinformatics 20(3): 340-348, 2003.
15. Duncan, J. W. and Steven, H. S., Collective dynamics of ‘small-world’ networks, Letters to Nature 393(6684): 440-2, 1998.
16. Watson, J. D. and Laskowski, R. A., Predicting protein function from sequence and structural data, Curr Opin Struct Biol 15(3): 275-284, 2005.
17. Goldberg, D. S. and Roth, F. P., Assessing experimentally derived interactions in a small world, Proc Natl Acad Sci U S A 100(8): 4372-4376, 2003.
18. Bu, D. and Zhao, Y., Topological structure analysis of the protein-protein interaction network in budding yeast, Nucleic Acids Res 31(9): 2443-50, 2003.
19. Lazaros K. G., Song, C. M., Havlin, S. and Herna, A. M., Scaling theory of transport in complex biological networks, Proceedings of National Academy of Sciences 104(19): 7746-7751, 2007.
20. Lu, H. and Zhu, X., The interactome as a tree--an attempt to visualize the protein-protein interaction network in yeast, Nucleic Acids Res 32(16): 4804-4811, 2004.
21. Tong, A. H. and Lesage, G., Global mapping of the yeast genetic interaction network, Science 303(5659): 808-813, 2004.
22. Wuchty, S. and Oltvai, Z. N., Evolutionary conservation of motif constituents in the yeast protein interaction network, Nat Genet 35(2): 176-179, 2003.
23. Tong, A. H. and Lesage, G., Global mapping of the yeast genetic interaction network, Science 303(5659): 808-813, 2004.
24. Mewes, H. W., Amid, C., MIPS: analysis and annotation of proteins from whole genomes, Nucleic Acids Res 32 Database issue: D41-D44, 2004.
25. Yu, J., and Fotouhi, F., Computational approaches for predicting protein-protein interactions: a survey, J Med Syst 30(1): 39-44, 2006.
26. Yu, X., Lu, L. J. and Mark, G., Integrated prediction of the helical membrane protein interactome in Yeast, J. Mol. Biol. 357(1): 339-49, 2006.