Biochemical and Biophysical Research Communications 448 (2014) 473–479

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Characterization of essential genes by topological properties in the perturbation sensitivity network

Lei Yang [a], Jizhe Wang [a], Huiping Wang [a], Yingli Lv [a], Yongchun Zuo [b,*], Wei Jiang [a,*]

[a] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, PR China
[b] The National Research Center for Animal Transgenic Biotechnology, Inner Mongolia University, Hohhot 010021, PR China

ARTICLE INFO

Article history:
Received 11 April 2014
Available online 4 May 2014

Keywords:
Yeast
Essential gene
Statistical test
Biological network

ABSTRACT

Genes that are indispensable for survival are called essential genes. In recent years, the analysis of essential genes has become extremely important for understanding the way a cell functions. With the advent of large-scale gene expression profiling technologies, it is now possible to profile transcriptional changes in the entire genome of *Saccharomyces cerevisiae*. Notwithstanding the accumulation of gene expression profiling in recent years, only a few studies have used these data to construct the network for *S. cerevisiae*. In this paper, based on the transcriptional profiling of the *S. cerevisiae* genome in hundreds of different gene disruptions, the perturbation sensitivity (PS) network is constructed. A scale-free topology with node degree following a power-law distribution is shown in the PS network. Twelve topological properties are used to investigate the characteristics of essential and non-essential genes in the PS network. Most of the properties are found to be statistically discriminative between essential and non-essential genes. In addition, the *F*-score is used to estimate the essentiality of each property, and the core number demonstrates the highest *F*-score among all properties.

1. Introduction

Essential genes are genes that are indispensable to support the survival of an organism [1]. These genes encode foundational functions required for a living cell under certain conditions and constitute a minimal gene set required for a living cell [2]. The deletion of only one of the essential genes is sufficient to result in infertility or lethality. Essential genes are involved in most survival-related housekeeping functions and tend to encode more hubs in the protein–protein interaction (PPI) network [3]. Some research results suggest that essential genes are correlated with disease genes [4], and essential genes of bacteria are attractive drug targets for new antibiotics. Therefore, the analysis of essential genes has been a major focus in genomic research and in drug design.

The behavior of a cell is a consequence of the complex interactions among its numerous constituents that constitute a series of complicated networks. The use of these networks to study biological systems has attracted emphasis among biologists in the last decade. However, most of the networks are too complex to be easily understood. Using graph theoretic concepts to investigate the topological properties of the networks, this problem can be overcome. The topological analysis in the networks has also become a useful tool for studying the social networks in social sciences [5], the characterization of drug-targets [6–8], the human disease genes [9,10], toxin-targets [11], and so on. With the advent of whole-genome expression profiling technologies, such as DNA microarrays, the transcriptional activity of thousands of genes in different biological conditions are simultaneously measured. Microarray experiments comparing expression levels of all genes in *Saccharomyces cerevisiae* for hundreds of mutants allow us to observe not only phenotypic changes of genes but also the up- or down-regulated genes in response to gene disruptions [12]. However, until recently, only a few studies have used these data to construct the network for *S. cerevisiae* [13,14].

In this study, the PS network is constructed from a transcriptional profiling study in response to 276 different gene disruptions in the yeast genome [12–14]. Nodes represent genes, and links are made between nodes if the expressions of the target genes are significantly altered by the disruption of the source gene. Twelve topological properties are calculated for each node in the PS network. Significant differences are found between the topological properties of essential genes and those properties of non-essential genes. We also employ the *F*-score to study the essentiality of each

* Corresponding authors. Fax: +86 471 5227683 (Y. Zuo). Fax: +86 451 8661 9617 (W. Jiang).
E-mail addresses: yczuo@imu.edu.cn (Y. Zuo), jiangwei@hrbmu.edu.cn (W. Jiang).

http://dx.doi.org/10.1016/j.bbrc.2014.04.136
0006-291X/

property and provide suggestions for a novel essential index for further study. The workflow of our study is shown in Fig. 1.

2. Materials and methods

2.1. Construction of the PS network

To construct a PS network from the genome-wide transcriptional profiling study for all genes in *S. cerevisiae*, 300 perturbation experiments, such as gene deletions or drug treatments, were downloaded from the supplemental data of Hughes et al. [12]. Among 300 perturbation experiments, 276 experiments were deletion mutants. Because drug treatments perturb more than one gene and could increase heterogeneity in the experiments, thus, drug treatments were not investigated in our study. We construct the PS network only based on the deletion mutants. In the work of Hughes et al., each pair of gene and deletion mutants was assigned a *P*-value according to the 'error model', correcting for gene measurement error and for biological noise [12]. A link is made between gene *i* and deletion mutant *j* if the expression of gene *i* is significantly changed ($P < 0.05$) in deletion mutant *j*. In our PS network, these links are represented as edges, and nodes are significantly regulated genes in deletion mutants or in the deleted genes. The *S. cerevisiae* PS network has 5638 genes as nodes, which are connected by 36841 edges.

2.2. Datasets of essential and non-essential genes

The lists of essential and non-essential genes of *S. cerevisiae* were obtained from the MIPS database [15] on February 20, 2014. The open reading frame (ORF) of an essential gene was regarded as an essential gene. In total, 949 essential genes and 4505 non-essential genes with unique ORF symbols were collected from the MIPS database. Among these genes, 849 essential genes were mapped into the PS network, and 4059 non-essential genes were mapped into the PS network. In total, 849 non-essential genes were randomly selected from 4059 non-essential genes. This dataset was defined as the control dataset.

Random sampling was also performed. In total, 849 non-essential genes were randomly selected from the 4059 non-essential genes. This procedure was repeated 1000 times to generate 1000 different balanced datasets containing different sets of non-essential genes.

2.3. Topological properties

In this study, the NetworkAnalyzer software [16] is used to calculate the degree, clustering coefficient [17], topological coefficient [18], average shortest path (ASP) and closeness centrality in the constructed network. The core number [19,20] is calculated using the MatlabBGL package, which is implemented in Matlab R2008a software. The Java plug-in cytoHubba [21] is used to explore important nodes or hubs in the network. This plug-in calculates four topological properties: betweenness, edge percolation component (EPC), maximum neighborhood component (MNC) and maximal clique centrality (MCC). The average distance to essential genes (ADEG) is defined as the average shortest distance between a gene and all essential genes in the PS network.

In this study, the common function index (CFI) [22] is defined to measure the amount of common Gene Ontology annotations of adjacent nodes in the PS network. The Gene Ontology annotations for *S. cerevisiae* were retrieved from the *Saccharomyces* Genome Database [23] on February 20, 2014. The CFI is defined as follows:

$$\mathrm{CFI}(i) = \sum_j d_{ij}^B + d_{ij}^c + d_{ij}^M$$

where *j* represents any node adjacent to node *i*. d_{ij}^B, d_{ij}^C and d_{ij}^M are the deepest ontology depth of common Gene Ontology annotations

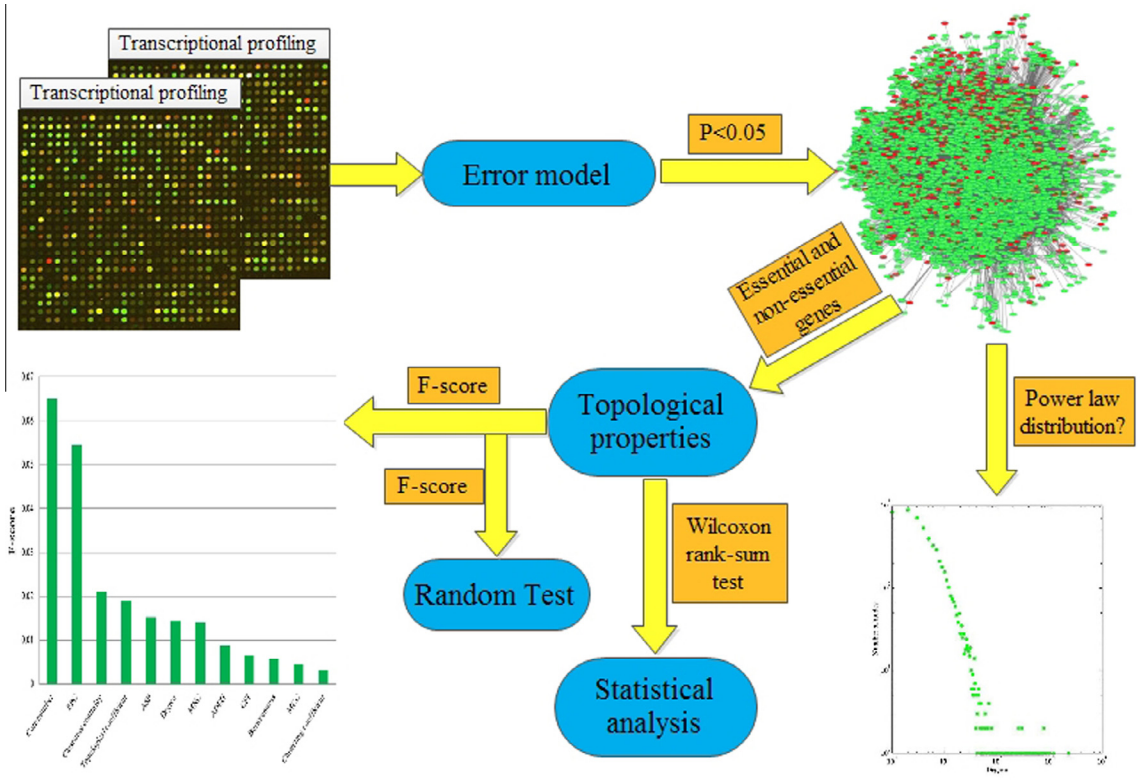

Fig. 1. The workflow of our study.

shared by i and j in the categories of biological process (B), cellular component (C) and molecular function (M), respectively.

2.4. Excess retention

The excess retention (ER) [20,24] of genes with property A (A indicates essential or non-essential) is defined as follows:

$$ER_k^A = (N_k^A/N_k)/(N^A/N)$$

where N_k is the total number of genes with the core number $\geqslant k$; N_k^A is the number of genes with a certain property, A, with the core number $\geqslant k$; N is the number of whole genes in the PS network; and N^A is the number of genes with a certain property, A, within the whole genes.

2.5. Statistical test

The Shapiro–Wilk (SW) test was used for testing the normality of the distributions. Because these variables did not follow a normal distribution in any of the categories, the Wilcoxon rank-sum test was used for assessing the statistical significance of the measures between different datasets in this study. All of the statistical analyses used in this work were performed using the freely available R package, version 3.0.2.

2.6. F-score

The F-score is a simple, intuitive method to measure the discrimination of two sets [25]. In this study, given training vectors x_k, $k = 1, \ldots, m$, the numbers of positive and negative instances are defined as n_+ and n_-, respectively, and then the F-score of the ith feature is defined as:

$$F1(i) = \left(\bar{x}_i^{(+)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(-)} - \bar{x}_i\right)^2$$

$$F2(i) = (1/(n_+ - 1))\sum_{k=1}^{n_+}\left(x_{k,i}^{(+)} - x_i^{(+)}\right)^2 + (1/(n_- - 1))\sum_{k=1}^{n_-}\left(x_{k,i}^{(-)} - x_i^{(-)}\right)^2$$

$$F\text{-score}(i) = F1(i)/F2(i)$$

where $\bar{x}_i$, $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$ are the averages of the ith feature of the whole, positive, and negative data sets, respectively; $\bar{x}_{k,i}^{(+)}$ is the ith feature of the kth positive instance; and $\bar{x}_{k,i}^{(-)}$ is the ith feature of the kth negative instance.

3. Results and discussion

3.1. Analysis of the PS network

The biological networks follow a power law distribution that corresponds to scale-free topology [26]. The scale-free topology is important for a reliable characterization of essentiality when the features of degree and betweenness are used [27]. To determine whether the PS network is consistent with scale free topology, the fitting of the node degree distribution to a power law is performed using the least squares method (Fig. 2A). As shown in Fig. 2A, the probability that a given yeast gene interacts with k other yeast genes follows a power law with an approximate R^2 coefficient of 0.721 and with the exponent of −1.059, a topology that is also shared by the PPI networks. This result indicates that the PS network forms a scale-free network and enables us to claim with more confidence for calculating topological properties in the network. A comprehensive set of topological properties of our PS network is calculated using the NetworkAnalyzer software to further address the properties of our network (Table 1). The network diameter is the largest distance between two nodes, and the PS network has the network diameter of 5. The network density and network centralization are also calculated. As shown in Table 1, the network density and centralization are 0.002 and 0.405, respectively.

3.2. Topological properties of essential genes in the PS network

To understand the topological properties of each essential gene in the PS network, 12 topological properties are calculated for each node. All of the results are shown in Table 2, Fig. S1 and in Fig. S2. Some previous works have shown that essential genes were more likely to be hubs in the PPI networks and that the most highly connected proteins in the cell were the most important proteins for its survival [22,28–30]. However, in our PS network, the results demonstrate that the degree of essential genes is significantly lower than that of non-essential genes ($P < 3.09\text{E}{-15}$) and that the average degree of non-essential genes is approximately quadruple as large as that in essential genes. Furthermore, to explore the important nodes in the PS network, the Hubba plug-in is used to analyze the network (Fig. 3, Figs. S3–S5). Using degree as an example, of the top 100 genes in the highest degree, 97 are non-essential genes (Fig. 3). These findings further suggest that high-degree nodes or hubs tend to be non-essential genes in the yeast PS network. In the PS network, if a gene is significantly up-or down regulated by more experiments of gene deletions or affected more genes in experiments of gene deletions, then this gene would have higher degree. In the work of Hughes et al. [12], only one essential gene was deleted in their experiments; thus, the degrees of essential genes primarily came from significant regulation in the gene deletion experiments. Because essential genes participate in more biological processes and have a longer evolutionary history, these genes might be more likely to be up-or down-regulated in the gene deletion experiments [14]. However, in our study, the lower degree of essential genes suggests that essential genes are robust to random perturbations and tolerant to outside transcriptional changes in numerous random perturbations, whereas non-essential genes are more liable to show transcriptional changes in a variety of environmental challenges. Based on the above results, we expect that essential genes are highly resistant to external perturbations, such as gene disruptions, through some types of buffering systems for maintaining the life of the organism.

The difference in the core number of the essential and non-essential genes is also investigated. The analysis of the core number can distinguish the topological role of highly connected nodes based on their locality, and a higher number represents the central placement of the subset in the original graph. In the PS network, there is significant difference in the core number between essential and non-essential genes ($P < 1.17\text{E}{-14}$). The violin plots indicate that each group shows an appreciable spread of core numbers (Fig. S2). The cumulative fraction distribution of the core number for non-essential genes stands out from that of essential genes (Fig. S1) because non-essential genes have a higher average core number than essential genes, which indicates that non-essential genes are more likely to be located at the backbone of the PS network. Furthermore, the excess retention of essential and non-essential genes is plotted (Fig. 2B). The excess retention of essential genes displays a marked decrease with the core number, whereas the excess retention of non-essential genes displays a marked increase with the core number. The analysis results presented in this study are similar to the results of Han et al. [13] and Ohn et al. [14]. This plot clearly demonstrates that essential genes are more likely to encode peripheral nodes in the PS network, whereas

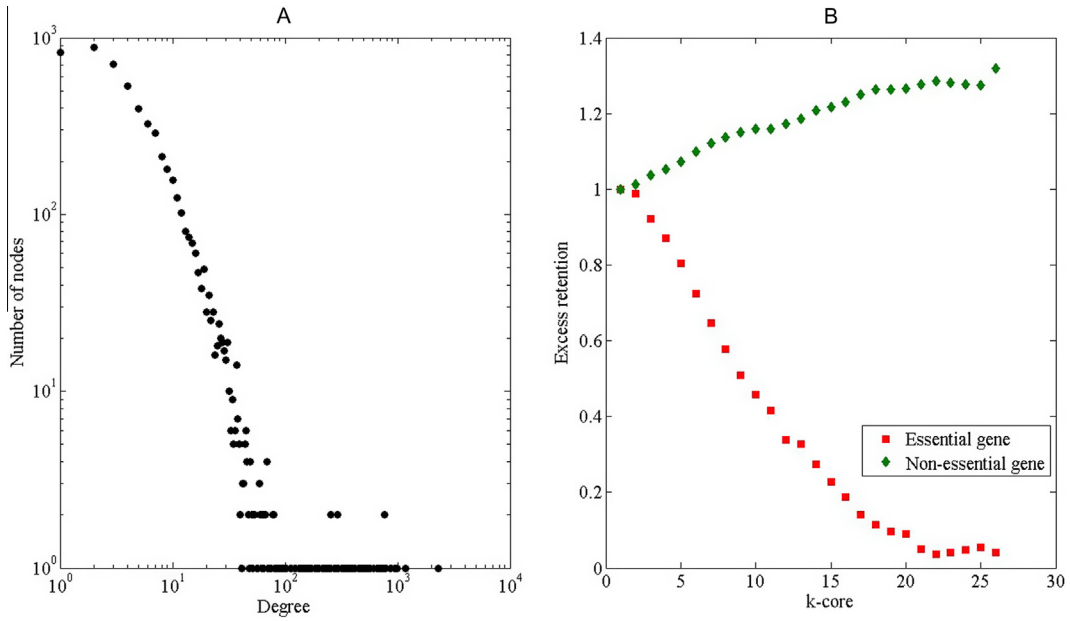

Fig. 2. (A) Degree distribution of the PS network. (B) Excess retention with the core number.

Table 1
Topological properties of the PS network.

Parameters		Parameters	
Number of nodes	5638	Betweenness	4902.527
Number of edges	36,841	Core number	6.649
ASP	2.739	EPC	56.060
Network diameter	5	Closeness centrality	0.371
Degree	12.871	Topological coefficient	0.264
Network density	0.002	Clustering coefficient	0.183
Network heterogeneity	4.784	MNC	9.611
Network centralization	0.405	MCC	114.062
ADEG	2.789	CFI	55.420

In this table, degree, ASP, ADEG, betweenness, core number, EPC, closeness centrality, topological coefficient, clustering coefficient, MNC, MCC and CFI are average values of all nodes in the network.

non-essential genes are more likely to encode hub nodes in the PS network.

The modularity of a node in the networks can be measured by the clustering coefficient. Inter-modular nodes tend to have a lower clustering coefficient, whereas close-connected module nodes tend to have higher clustering coefficient. In the PS network, the average clustering coefficient of an essential gene is higher than that of a non-essential gene; however, the difference between these genes is not significant ($P < 6.66E{-}1$). The difference in the topological coefficients of the two gene groups is also investigated. In contrast to the results of the clustering coefficient, there is a significant group difference between the topological coefficients ($P < 3.42E{-}10$). The average topological coefficients of essential and non-essential genes are 0.2988 and 0.2489, respectively, which indicates that essential genes have on average 1.2 times more topological coefficients than non-essential genes in the network.

Compared with the control dataset, the average closeness centrality is significantly lower in the essential gene dataset ($P < 8.30E{-}9$), indicating that non-essential genes have fast information spreads from a given node to all other reachable nodes in the PS network. Betweenness can be regarded as a measure of the influence a node has over the spread of information through the network. The nodes with larger betweenness represent the bottlenecks in the network or pathways in which these nodes participate. The average betweenness of essential genes is significantly lower than that of non-essential genes ($P < 1.51E{-}11$). The ASP of essential genes is significantly longer than that of non-essential genes, indicating that essential genes communicate slowly in the PS network. In the PS network, ADEG is also statistically discernible between essential and non-essential genes ($P < 1.47E{-}5$). The

Table 2
The topological properties between essential and non-essential genes.

Property	Essential gene		Non-essential gene		*P*-value
	Mean	SD	Mean	SD	
Degree	4.6313	3.8359	17.1270	73.7330	3.08E−15
Core number	4.5960	3.7392	7.4323	6.9187	1.16E−14
Clustering coefficient	0.2000	0.2557	0.1733	0.2099	6.65E−01
Topological coefficient	0.2988	0.1838	0.2489	0.1780	3.41E−10
Closeness centrality	0.3630	0.0397	0.3755	0.0464	8.29E−09
Betweenness	255.25	407.98	7678.2	69237	1.50E−11
ASP	2.7898	0.3232	2.7063	0.3547	8.29E−09
ADEG	2.8214	0.3376	2.7560	0.3624	1.46E−05
CFI	33.5241	39.8920	66.1991	281.6982	6.20E−04
EPC	38.6110	29.2250	63.9940	71.1140	4.94E−14
MCC	10.9390	30.5520	230.0700	2292.9	4.26E−10
MNC	3.1861	3.3703	13.0920	58.9080	9.22E−13

In this table, SD indicates standard deviation.

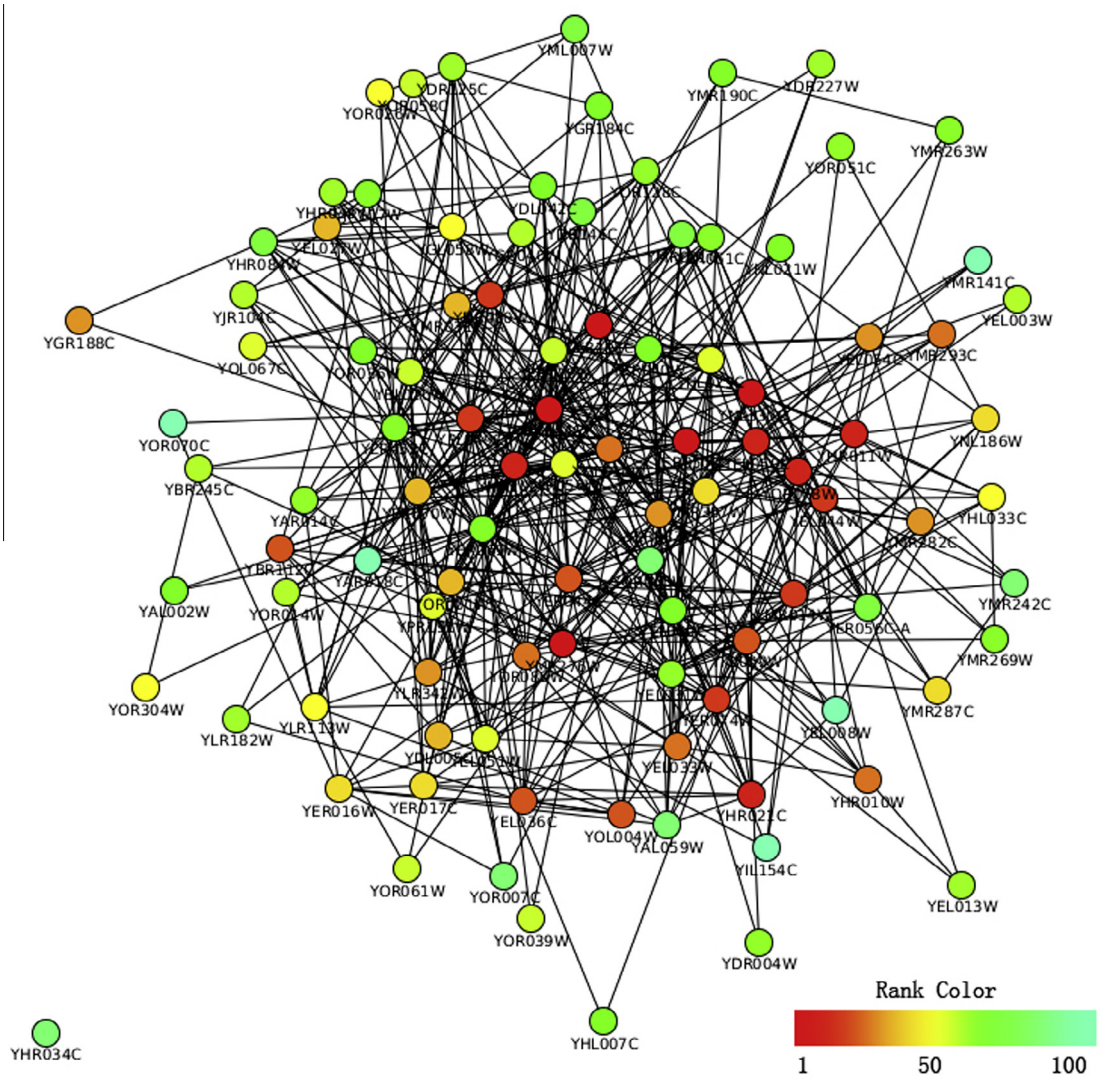

Fig. 3. The subnetwork that is based on the top 100 nodes in the highest degree. All the non-essential genes are used in this case.

higher ADEG indicates that essential genes have longer average distances to other essential genes in the PS network. CFI is used to measure the amount of common functions of adjacent nodes in the PS network, and there is a significant difference in CFI between essential and non-essential genes.

In this work, EPC, MCC and MNC, which are calculated by cytoHubba, are used for exploring important nodes or hubs in our network. There are significant differences in EPC, MCC and MNC between essential and non-essential genes ($P < 4.95E{-}14$ for EPC, $P < 4.27E{-}10$ for MCC and $P < 9.23E{-}13$ for MNC), and non-essential genes have higher average values than essential genes. These results further suggest that non-essential genes are more likely to be hubs in the PS network. These results are expected because the degree of non-essential genes is higher than that of essential genes in our PS network.

3.3. Random sampling

To avoid statistical bias, random sampling is also performed to characterize the properties of essential and non-essential genes. The distributions of the mean property levels of 1000 samples and *P*-value levels of 1000 samples are shown in Fig. S6 and in Fig. S7. There are significant differences in the degree, core number, topological coefficient, ASP, betweenness, closeness centrality, EPC, MCC and MNC between essential and non-essential genes, which are consistent with the above results. However, there are no significant differences in the clustering coefficients between essential and non-essential genes. For ADEG and CFI, in most cases, the differences between essential and non-essential genes are significant.

3.4. Features for gene essentiality

The discrimination of two different features can be roughly evaluated by the *F*-score. The larger the *F*-score is, the more likely discriminative this feature is. Thus, the *F*-score can be used to evaluate the essentiality of the features (Fig. 4).

The core number has the most discriminative feature, whereas the clustering coefficient has the least discriminative feature. The lowest *F*-score of the clustering coefficient is expected because there is no significant difference in the clustering coefficients between essential and non-essential genes. In contrast, the highest *F*-score of the core number is more surprising. One of the reasons why such feature is more discriminative might be due to the consequence of the structure of the PS network. As shown in the previous section, non-essential genes are more likely to be located at the backbone of the network compared with other locations in

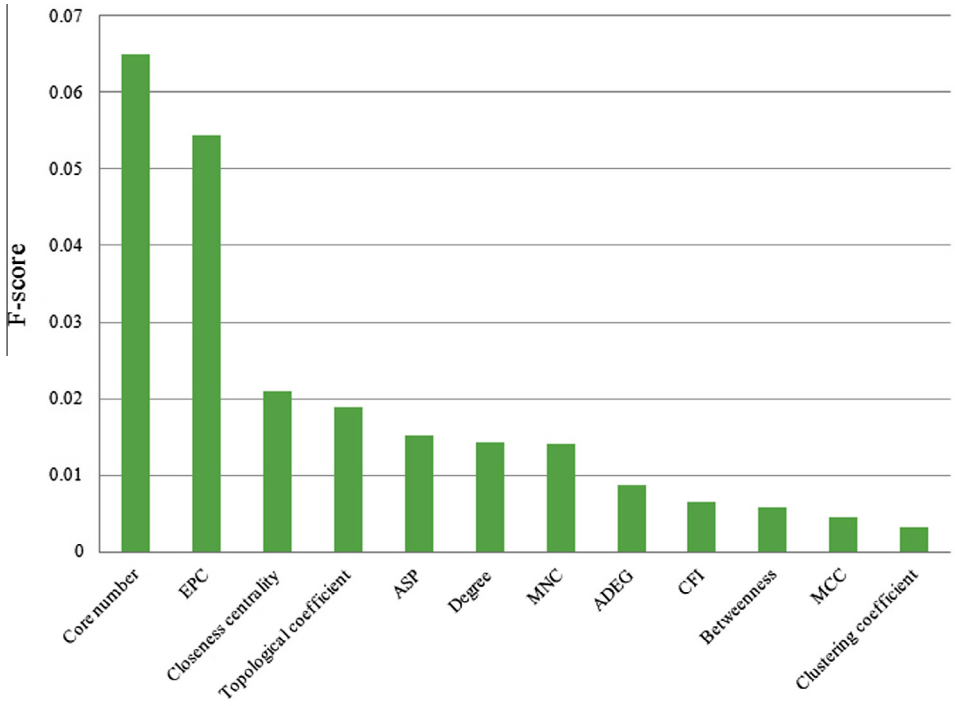

Fig. 4. The *F*-scores of 12 topological properties.

the PS network. Previous studies have discussed the relation between the degree and essentiality of genes in the PPI networks [7,18,30]. These studies found that proteins with a high degree are more likely to be essential proteins in the PPI networks, although the relation between a gene's essentiality and degree is not deterministic or simple. Although the PS network is different from the commonly used PPI networks in other studies, we would expect that the *F*-score of degree should be higher in the PS network. As expected, the *F*-score of the degree is 0.01432, which is ranked sixth in all features. However, the core number, EPC, closeness centrality, topological coefficient and ASP tend to have higher *F*-scores than the degree, indicating that these features may be more effective for reflecting the essentiality of genes than the degree in the PS network.

Random sampling is also performed to reduce the statistical bias of our study. All of the results are shown in Fig. S8 and in Dataset S1. In 1000 random samplings, the *F*-scores of the core number and EPC are always higher than that of the degree, indicating that these topological properties may be more effective for reflecting the essentiality of genes than the degree. However, we should also keep in mind that the essentiality index of each property is roughly estimated by the *F*-score in this study and that the reliability and usefulness of the *F*-score for reflecting the essentiality index of each property must be proven in future work.

4. Conclusions

In this study, we have built the PS network of *S. cerevisiae* using the mRNA expression profiles of 276 gene deletion experiments. The PS network follows a power law distribution that corresponds to scale-free topology. Twelve topological properties are used to assess the differences in topological properties between essential and non-essential genes. Except for the clustering coefficient, ADEG and CFI, all topological properties are significantly different between essential and non-essential genes. Essential genes tend to encode peripheral nodes, whereas non-essential genes tend to encode hub nodes in the PS network. Furthermore, the properties for the essentiality of genes are also evaluated by the *F*-score, and the core number is found to be the most discriminative index in all properties. We hope that these findings will contribute to elucidating the function of essential genes and provide useful help for the ever growing field of related areas. However, some limitations exist in our study that suggest avenues for future work. We only examined *S. cerevisiae* mRNA profiles. We suspect that the same trends will be detectable in other species. We anticipate that the same work will become important future topics in genomics.

Acknowledgments

This work was supported by the Scientific Research Fund of Heilongjiang Provincial Education Department (No. 12521303).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.04.136.

References

[1] M. Seringhaus, A. Paccanaro, A. Borneman, M. Snyder, M. Gerstein, Predicting essential genes in fungal genomes, Genome Res. 16 (2006) 1126–1135.
[2] M. Juhas, L. Eberl, J.I. Glass, Essence of life: essential genes of minimal genomes, Trends Cell Biol. 21 (2011) 562–568.
[3] M.L. Acencio, N. Lemke, Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information, BMC Bioinformatics 10 (2009) 290.
[4] S.J. Furney, M.M. Albà, N. López-Bigas, Differences in the evolutionary history of disease genes affected by dominant or recessive mutations, BMC Genomics 7 (2006) 165.
[5] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994.
[6] M. Zhu, L. Gao, X. Li, Z. Liu, C. Xu, Y. Yan, E. Walker, W. Jiang, B. Su, X. Chen, The analysis of the drug-targets based on the topological properties in the human protein–protein interaction network, J. Drug Target. 17 (2009) 524–532.
[7] M.A. Yıldırım, K.I. Goh, M.E. Cusick, A.L. Barabási, M. Vidal, Drug-target network, Nat. Biotechnol. 25 (2007) 1119–1126.
[8] C. Wang, W. Jiang, W. Li, B. Lian, X. Chen, L. Hua, H. Lin, D. Li, X. Li, Z. Liu, Topological properties of the drug targets regulated by microRNA in human protein–protein interaction network, J. Drug Target. 19 (2011) 354–364.

[9] K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, A.L. Barabasi, The human disease network, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 8685–8690.
[10] J. Xu, Y. Li, Discovering disease-genes by topological features in human protein–protein interaction network, Bioinformatics 22 (2006) 2800–2805.
[11] L. Yang, J. Wang, H. Wang, Y. Lv, Y. Zuo, W. Jiang, Analysis and identification of toxin targets by topological properties in protein–protein interaction network, J. Theor. Biol. 349 (2014) 82–91.
[12] T.R. Hughes, M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, H. Dai, Y.D. He, Functional discovery via a compendium of expression profiles, Cell 102 (2000) 109–126.
[13] H.W. Han, J.H. Ohn, J. Moon, J.H. Kim, Yin and Yang of disease genes and death genes between reciprocally scale-free biological networks, Nucleic Acids Res. 41 (2013) 9209–9217.
[14] J.H. Ohn, J. Kim, J.H. Kim, Genomic characterization of perturbation sensitivity, Bioinformatics 23 (2007) i354–i358.
[15] H.W. Mewes, D. Frishman, U. Güldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Münsterkötter, S. Rudd, B. Weil, MIPS: a database for genomes and protein sequences, Nucleic Acids Res. 30 (2002) 31–34.
[16] Y. Assenov, F. Ramírez, S.E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2008) 282–284.
[17] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[18] D.S. Goldberg, F.P. Roth, Assessing experimentally derived interactions in a small world, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 4372–4376.
[19] S. Wachi, K. Yoneda, R. Wu, Interactome–transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues, Bioinformatics 21 (2005) 4205–4208.
[20] S. Wuchty, E. Almaas, Peeling the yeast protein network, Proteomics 5 (2005) 444–449.
[21] C.Y. Lin, C.H. Chin, H.H. Wu, S.H. Chen, C.W. Ho, M.T. Ko, Hubba: hub objects analyzer a framework of interactome hubs identification for network biology, Nucleic Acids Res. 36 (2008) W438–W443.
[22] Y.C. Hwang, C.C. Lin, J.Y. Chang, H. Mori, H.F. Juan, H.C. Huang, Predicting essential genes based on network and sequence analysis, Mol. BioSyst. 5 (2009) 1672–1678.
[23] J.M. Cherry, C. Adler, C. Ball, S.A. Chervitz, S.S. Dwight, E.T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, SGD: Saccharomyces genome database, Nucleic Acids Res. 26 (1998) 73–79.
[24] S. Wuchty, Evolution and topology in the yeast protein interaction network, Genome Res. 14 (2004) 1310–1314.
[25] Y.W. Chen, C.J. Lin, Combining SVMs with various feature selection strategies, in: Feature Extraction, Springer, 2006, pp. 315–324.
[26] A.L. Barabasi, Z.N. Oltvai, Network biology: understanding the cell's functional organization, Nat. Rev. Genet. 5 (2004) 101–113.
[27] A.F. Flórez, D. Park, J. Bhak, B.C. Kim, A. Kuchinsky, J.H. Morris, J. Espinosa, C. Muskus, Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection, BMC Bioinformatics 11 (2010) 484.
[28] H. Yu, D. Greenbaum, H. Xin, M. Zhu, X. Lu, Gerstein, genomic analysis of essentiality within protein networks, Trends Genet. 20 (2004) 227–231.
[29] K. Pang, H. Sheng, X. Ma, Understanding gene essentiality by finely characterizing hubs in the yeast protein interaction network, Biochem. Biophys. Res. Commun. 401 (2010) 112–116.
[30] H. Jeong, S.P. Mason, A.L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.