Identifying essential proteins based on sub-network partition and prioritization by integrating subcellular localization information

Min Li, Wenkai Li, Fang-Xiang Wu, Yi Pan, Jianxin Wang

PII: S0022-5193(18)30143-7
DOI: 10.1016/j.jtbi.2018.03.029
Reference: YJTBI 9406

To appear in: *Journal of Theoretical Biology*

Received date: 12 August 2017
Revised date: 19 March 2018
Accepted date: 20 March 2018

Highlights

- Proposing essential protein prediction method SPP by integrating PPI network and subcellular localization. data
- SPP achieves higher prediction accuracy compared with existing computational methods on YDIP and YBioGRID network.
- Sub-network partition and prioritization can effectively reduce the effect of false positives in PPI networks.

Identifying essential proteins based on sub-network partition and prioritization by integrating subcellular localization information

Min Li[a,*], Wenkai Li[a], Fang-Xiang Wu[b], Yi Pan[c], Jianxin Wang[a,*]

[a] *School of Information Science and Engineering, Central South University, Changsha, 410083, China*
[b] *Division of Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada*
[c] *Department of Computer Science, Georgia State University, Atlanta, GA 30302-4110, USA*

Abstract

Essential proteins are important participants in various life activities and play a vital role in the survival and reproduction of living organisms. Identification of essential proteins from protein-protein interaction (PPI) networks has great significance to facilitate the study of human complex diseases, the design of drugs and the development of bioinformatics and computational science. Studies have shown that highly connected proteins in a PPI network tend to be essential. A series of computational methods have been proposed to identify essential proteins by analyzing topological structures of PPI networks. However, the high noise in the PPI data can degrade the accuracy of essential protein prediction. Moreover, proteins must be located in the appropriate subcellular localization to perform their functions, and only when the proteins are located in the same subcellular localization, it is possible that they can interact with each other. In this paper, we propose a new network-based essential protein discovery method based on sub-network partition and prioritization by integrating subcellular localization information, named SPP. The proposed method SPP was tested on

[*] Corresponding author
Email addresses: limin@mail.csu.edu.cn (Min Li), lwktechnology@csu.edu.cn (Wenkai Li), faw341@mail.usask.ca (Fang-Xiang Wu), yipan@gsu.edu (Yi Pan), jxwang@csu.edu.cn (Jianxin Wang)

two different yeast PPI networks obtained from DIP database and BioGRID database. The experimental results show that SPP can effectively reduce the effect of false positives in PPI networks and predict essential proteins more accurately compared with other existing computational methods DC, BC, CC, SC, EC, IC, NC.

Keywords: essential protein, protein-protein interaction network, network topology, biological information

1. Introduction

Proteins are the basic substances for the survival of living organisms, and they are also the fundamental substances constituting biological cell and tissue and maintaining life activities [1, 2]. In addition, proteins are indispensable to physiological functions and closely related to the physiological states within living organisms. There are many types of proteins in living organisms, which have different biological functions and join in various living processes, such as nutrient transport, immune response, biochemical reaction, and so on. The deletion of some proteins results in the demise of organisms, and leads to disease or affects the growth. Such proteins are called essential proteins [3]. It is very important to identify essential proteins, which contributes to understanding the minimum requirements of the survival and development of a cell [4], analyzing the causes of disease [5] and discovering the mechanisms of drug actions.

Traditional experiment methods, such as single gene knockout [6], RNA interference [7] and conditional knockout [8], have been used to detect essential proteins. However, these experiment methods cannot meet the needs of proteomics research because of the high cost and long period to obtain experimental results. With the development of high-throughput technologies, including yeast two-hybrid system [9], mass spectrometry analysis [10], protein chips [11], tandem affinity purification [12], there is increasing data produced. As a result, several PPI databases have been set up, such as Database of Interacting Proteins (DIP) [13], the Biological General Repository for Interaction Dataset

(BioGRID) [14], the Molecular Interaction database (MINT) [15], the Biomolecular Interaction Network Database (BIND) [16], etc. For now, a number of computational methods have been presented for discovering essential proteins from PPI networks [17, 18, 19].

It has been shown by Jeong et al. [20] that the most highly connected proteins in the cell have a vital function for its survival, also called centrality-lethality rule, through investigating the yeast PPI network. The absence of highly connected protein nodes in the PPI network can lead to the collapse of the whole network structure and the fatal effect on the living organism itself. Centrality-lethality rule has also been applied on nematode, fly and other species, and was found to work well on these species [21, 22, 23]. With more and more studies on network-based computational methods for identifying essential proteins, there are a series of centrality methods that came out and widely applied. According to the different ways to measure the essentiality of each protein in PPI network, they generally can be divided into two groups: one is local-based connection methods, apart from the above degree centrality(DC), it also includes sub-graph centrality (SC) [24], eigenvector centrality (EC) [25], maximum neighborhood component (MNC) [26], local average connectivity (LAC) [27], network centrality (NC) [28], topology potential-based method (TP) [29], etc. and the other is global-based connection methods, such as betweenness centrality (BC) [30], closeness centrality (CC) [31], information centrality (IC) [32], bottle neck (BN) [33], etc. And most of the above network centralities can be calculated by the network analysis tools, such as CytoNCA [34] or DyNetViewer [35].

For network-based computational methods, the reliability of the PPI network can have an effect on the accuracy for identifying essential proteins. In order to solve this problem effectively, researchers have taken several measures. The most popular are the three following approaches: constructing weighted network [36, 37, 38], filtering noise data [39, 40, 41], or integrating other biological data [42, 43, 44]. For the first one, researchers generally constructed weighted networks by using different methods and calculated the proteins essentiality based on the weighted network. Considering that non-essential proteins may have lots of interactions but their neighbors have few interactions, Cheng et al. [36] proposed a novel computational method based on the local neighborhood connectedness from weighted PPI networks, named LNCw. Tang et al. [37] presented the weighted degree centrality through computing Pearson correlation coefficient (PCC) and edge clustering coefficient to weight each interaction in PPI networks. In our previous studies, we constructed weighted PPI networks by combining the logistic regression-based model and the similarity of protein function, and redefined six standard centrality measures for ranking proteins in weighted networks [38]. The second one is to filter noise. There are different refining methods developed, such as LSED [39] and TS-PIN [40]. The third one is to integrate multivariate data, such as subcellular localization information [42, 43, 44], orthology [42, 45], priori knowledge [19], gene expression data [46, 47] and protein complexes [48, 49], etc. Among them, the subcellular localization information is easily gathered, and has been widely applied in the identification of essential proteins [42, 43], protein complexes [50, 51], and the prediction of protein function [52, 53, 54].

In this paper, we propose a new network-based essential protein prediction method, named SPP, based on sub-network partition and prioritization by integrating subcellular localization information. The basic idea of this method is that one protein can have an interaction with another protein only if they exist in the same subcellular compartments [55], so the original PPI network can be divided into several sub-networks, which can effectively reduce the influence of noise and promote the reliability of PPI network. To evaluate the performance of essential protein prediction, we used two yeast PPI networks from different sources and compared with seven existing methods, including DC, BC, CC, SC, EC, IC, and NC. The PPI networks obtained from the DIP database and the BioGRID database, in which the self-interactions and duplicate interactions are removed. The experimental results on the identification of essential proteins have shown that our method outperforms other network-based computational method.

2. Materials and Methods

The existing network-based methods identify essential proteins by investigating the topological characters of essential proteins from different views. Though great progresses have been achieved, the predicted precision is limited because of the high noise in the high-throughput PPI data. Peng et al. [39] used subcellular localization information to recheck the centrality-lethality rule. In the previous studies, it has been shown that subcellular localization information is useful to reduce the false positives in the PPI data [56] and contributes to the accuracy of the essential proteins prediction. Compared with gene expression data and orthologous information, the subcellular localization information is more easily obtained. In this study, we propose a new network-based essential protein discovery method SPP based on sub-network partition and prioritization by integrating subcellular localization information

2.1. Experimental Dataset

Yeast is currently the widely used species for studying essential proteins. In this study, we obtained two different sizes of networks from DIP and BioGRID, respectively. The yeast network from DIP is named YDIP, which is composed of 4746 proteins and 15166 interactions after self-interactions and duplicate interactions were removed. Another one from BioGRID is named YBioGRID, which is much denser than YDIP. There are 52833 interactions connecting 5616 proteins in YBioGRID. A set of known essential proteins are collected from DEG [57], MIPS [58], SGD [59] and SGDP [60]. After preprocessing, there are 1130 essential proteins in YDIP and 1199 essential proteins in YBioGRID, the detailed information of YDIP and YBioGRID is shown in Table 1. The yeast subcellular localization data is downloaded from COMPARTMENTS database [61]. There are eleven localization categories: cytoskeleton (CN), cytosol (CL), endoplasmic (EC), endosome (EE), extracellular (ER), golgi (GI), lysosome (LE), mitochondrion (MN), nucleus (NS), peroxisome (PE) and plasma (PA).

Table 1: **Detailed information of the PPI networks**

Dataset	Proteins	Interactions	Average degree	Essential proteins
YDIP	4746	15166	6.39	1130
YBioGRID	5616	52833	18.82	1199

2.2. Sub-network partition and prioritization

As mentioned by Briesemeister [62], based on the different roles of proteins, they would be transported to different subcellular localizations after being synthesized, some may be involved in more than one subcellular localization. Kumar et al. [52] thought most of proteins take part in maintaining the integrity of cellular structure. They also gave an example that 34% of proteins which are located in nuclear, participate in the process of transcription, and 26% of mitochondrial-localized proteins act on cellular respiration. Researches on protein subcellular localization have revealed that the localization of proteins within cellular microenvironments has a strong correlation between functions [63, 64] and essentiality [39, 45] of a protein.

To figure out the number of proteins and essential proteins located in different subcellular localizations in yeast PPI network, we obtained two different sizes of networks from DIP and BioGRID, respectively. The yeast network from DIP is named YDIP, which is composed of 4746 proteins and 15166 interactions after self-interactions and duplicate interactions were removed. Another one from BioGRID is named YBioGRID, which is much denser than YDIP. There are 52833 interactions connecting 5616 proteins in YBioGRID. The analysis results are shown in Figure 1.

From Figure 1, we can see that, the majority of either whether the proteins or the essential proteins in the network are mainly distributed in NS, MN, CL and EC within YDIP, as well as YBioGRID. Through using subcellular localization information, we think it may be beneficial for identifying more essential proteins by partitioning the overall PPI network into several sub-networks, in which proteins are located in the same cellular compartment. It is clear to know how many proteins are in each sub-network while not knowing how many essential proteins are. Therefore, we determine the priority of sub-networks by their scales, i.e. the number of nodes in the network. The higher the priority, the higher the importance of this sub-network is, which provides an important basis to estimate the essentiality of nodes within the network in the next step. As shown in Figure 1(a), the numbers of proteins on YDIP in different subcellular localizations (NS, MN, CL, EC, CN, PA, GI, EE ,PE, ER, LE) are 1781, 448, 363, 242, 178, 166, 131, 81, 31, 3, 2, respectively. It indicates that NS contains the highest priority sub-network, followed by MN, and then CL, EC, CN, PA, GI, EE, PE, ER, LE.

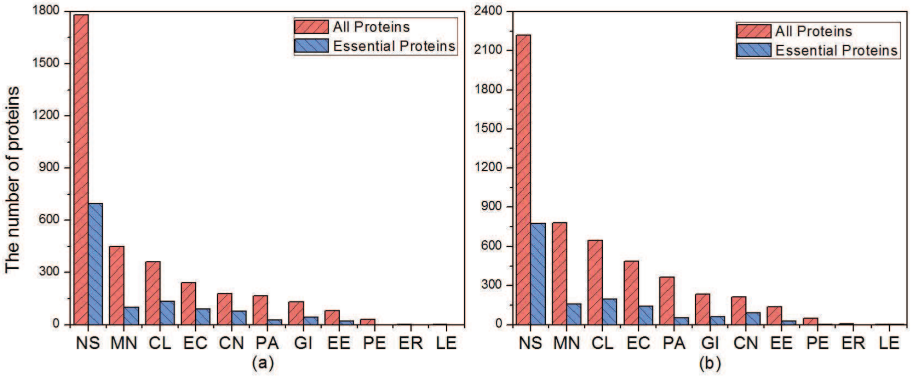


Figure 1: The number of proteins and essential proteins in different subcellular localizations on yeast PPI network. (a) YDIP; (b) YBioGRID

A PPI network can be represented as an undirected graph $G = (V, E)$ with a node $v \in V$ representing a protein and an edge $(u, v) \in E$ representing an interaction between protein u and protein v. The sub-network for subcellular location i is denoted as $G^i = (V^i, E^i)$. For example in Figure 2, a protein node $v(v \in V)$ is located in four cellular compartments (NS, MN, PA, ER) and has nine neighbor nodes (a, b, c, d, e, f, g, h, i). Here, we use G_v to denote the graph composed by these nodes. Among them, the protein node v and its two neighbors (a, b) are in the same cellular compartment (PA). Similarly, because node b exists in multiple subcellular compartments, node v and b with the other three nodes (c, d, e) locate in the same compartment (NS). Finally, we can partition the network G_v into four sub-networks, i.e. $G^{PA}, G^{NS}, G^{MN}, G^{ER}$, and the priority order is $G^{NS} > G^{MN} > G^{PA} > G^{ER}$.

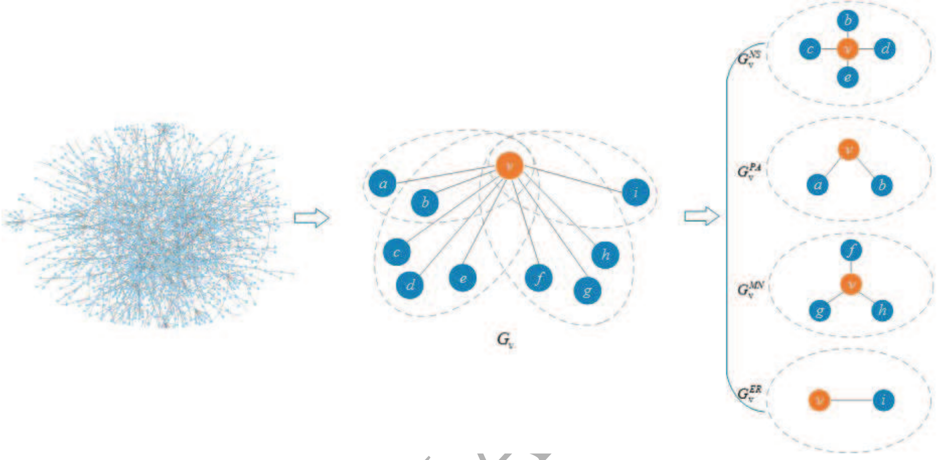


Figure 2: Example of the division of PPI network based on subcellular localization information

2.3. Method SPP

After sub-network partitioning, it is important to evaluate the essentiality of proteins. Centrality-lethality rule illustrates that the most highly connected proteins have a vital function in the cell [20]. According to this theory, degree centrality for identifying essential proteins is proposed to evaluate the important of nodes by their degree. In our previous work, through comparing all the nodes in the PPI network with known essential proteins, we found only 49.8% of proteins are essential proteins out of 305 protein nodes with a degree more than 20, and believed that neighbors of non-essential proteins which interact with much proteins have less interactions, even no interaction with other proteins [27]. This indicates that only using node degree may be not enough to identify essential proteins effectively. After studying on the interaction of yeast proteins, Pereira-Leal et al. [65] discovered that the essential-essential interactions can form a giant component in consideration of the preference connecting between essential proteins. In addition, Butland et al. [66] proposed that the essential 175 proteins are more conserved, and these highly conserved proteins are highly connected and construct a stable interaction network. To validate the current dataset, we visualize analysis on the YDIP and depict one highly connected component of essential proteins. From Figure 3, the non-essential proteins are renamed with the beginning of the NON string to clearly distinguish from the essentials.

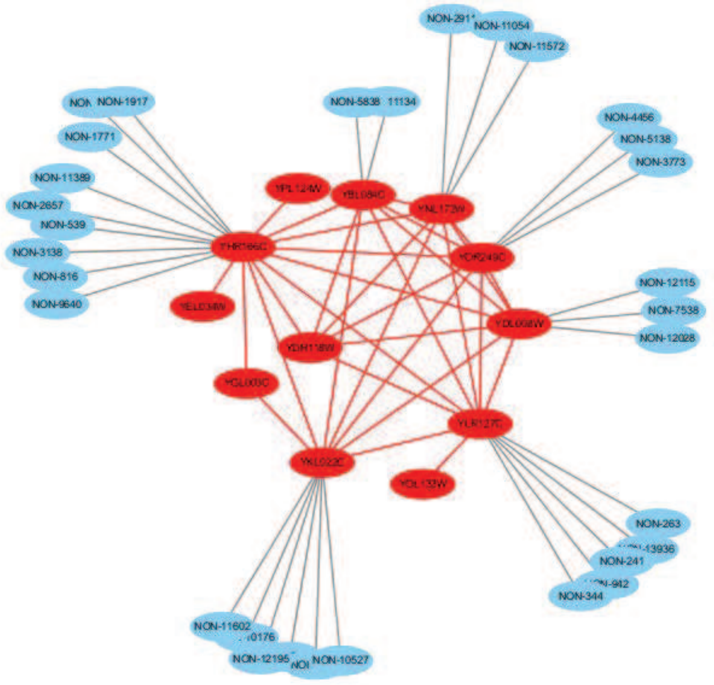


Figure 3: A connected component from YDIP. Red nodes denotes essential proteins, blue nodes represent non-essential proteins

180 In each sub-network, we use common neighbors between two nodes to measure the strength of their connection, represented as SCN_i, for subcellular localization i. Specifically, for the two nodes v and u, the number of their common

neighbors can be defined as $SCN_i(v,u)$.

$$SCN_i(v,u) = |N_i(v) \cap N_i(u)| \quad v,u \in V^i \tag{1}$$

where $N_i(v)$ and $N_i(u)$ represent the set of neighbors of node v and node u in sub-network G^i, respectively. $|N_i(v) \cap N_i(u)|$ denotes the number of common neighbors of node v and node u . As shown in Figure 4, the internal structure of sub-network G_v^{NS}. We can see that node v has two common neighbors (c, d) with node e, so $SCN_{NS}(v,e) = 2$, and $SCN_{NS}(v,c) = 1$. Through the analysis above, we can calculate the comprehensive score of each node in the sub-network i. The score of protein v sub-network i is defined as $SPP_i(v)$:

$$SPP_i(v) = \sum_{u \in N(v)} \frac{Max(d_v,d_u)}{d_v + d_u} \times SCN_i(v,u) \tag{2}$$

where d_v denotes the degree of node v in the sub-network, and the entire fraction is an adjustment parameter to highlight the influence of high-connectivity nodes compared with low-connectivity nodes.

Eventually, we use a dual-priority to rank all the proteins in a given PPI network. We firstly choose the sub-network with a high priority in turn and sort all its proteins by the SPP value in descending order. Moreover, because one protein maybe exists in more than one subcellular localization, we only analyze the sub-network with the highest priority. In general, the higher the sub-network priority and the higher the score of a protein, the more likely the protein is an essential protein

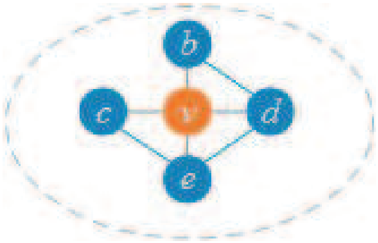


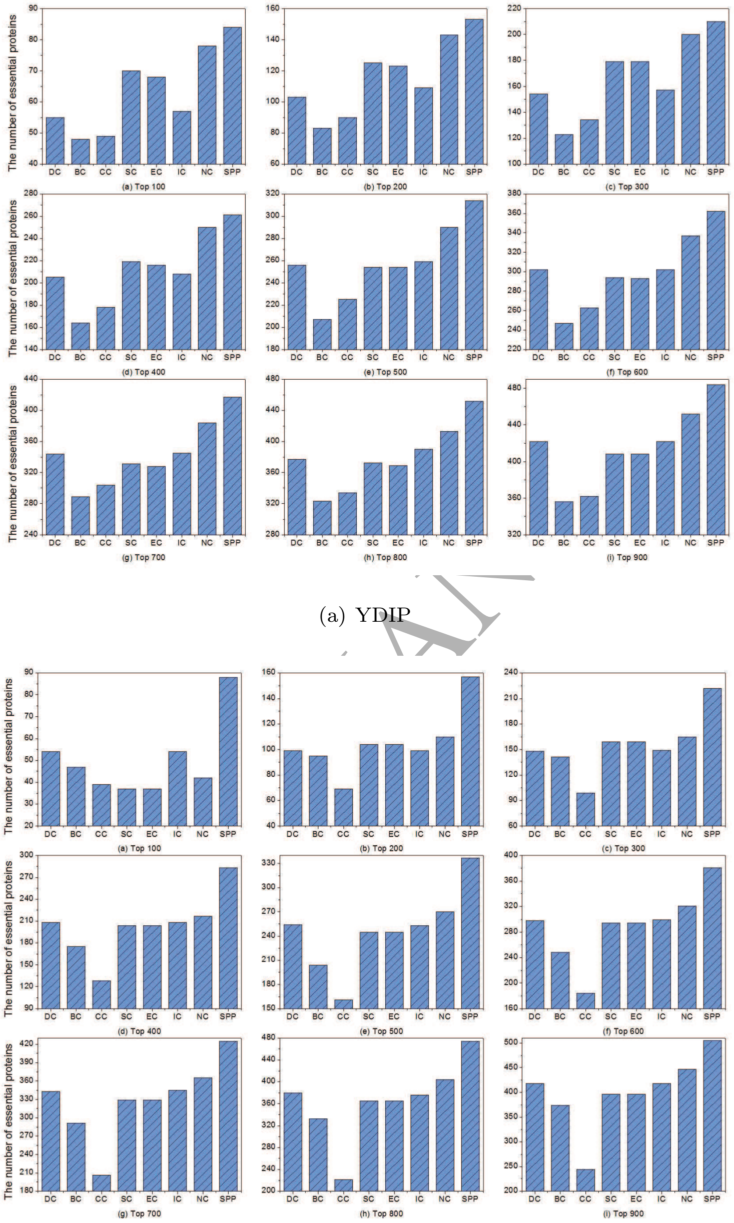
Figure 4: Example of a sub-network G_v^{NS}

3. Results

In this study, we compared SPP with other seven network-based essential protein prediction methods, including DC, BC, CC, SC, EC, IC and NC. We can directly rank proteins by the values of calculation results for competing methods. However, for SPP we should get sub-networks in turn according to the order of the sub-network prioritization and rank proteins based on the score of SPP in this sub-network, so one protein can be at the top of all proteins only when it is located in the sub-network with the highest priority and has the maximum score. Based on the rule of ranking and screening, we sequentially select a certain number of proteins to construct a candidate set of predicted essential proteins. By comparing with the known essential proteins, we can obtain statistics about the number of real essential proteins within the candidate set.

3.1. Comparison with seven representative methods

In order to evaluate the effectiveness of the proposed method, we select the top 100, 200, 300, 400, 500, 600, 700, 800 and 900 proteins as the candidate set. The number of essential proteins identified by SPP and other seven prediction methods are shown in Figure 5. As can be seen from this figure, the performance to identify essential proteins by using SPP has been greatly improved on both YDIP and YBioGRID. From Figure 5(a), we can see that identification accuracy of SPP achieves 84%, 76.5%, 70%, 65.2%, 62.8%, 60.3%, 59.6%, 56.5% and 53.8% with different top set on YDIP. Comparing with DC, which is a simple and widely used centrality method, the results are raised 52.7%, 48.5%, 36.4%, 27.3%, 22.7%, 19.9%, 21.2%, 19.9% and 14.7% for the nine candidate sets, respectively. Especially, as the best one of the seven network-based methods, the prediction results of NC is also improved by 7.7%, 6.9%, 8.6% and 9.4% in top 100, 200, 700, 800 essential candidates, respectively. The same result can be obtained from Figure 5(b). Even compared with DC and NC, experimental result is improved by 62.9% and 137% in top 100 proteins on YBioGRID, respectively.



(a) YDIP

(b) YBioGRID

Figure 5: Comparison of the number of essential proteins detected by SPP and other methods (DC, BC, CC, SC, EC, IC, and NC) from yeast PPI network. (a) YDIP; (b) YBioGRID

3.2. Validated by jackknifing methodology

In this section, we evaluate the performance of SPP and seven other prediction methods with the jackknifing methodology. The experimental results can be seen in Figure 6. The x axis is the number of top ranked proteins, and y axis represents the cumulative number of essential proteins identified by prediction methods. From this figure, we can clearly see that SPP has the best performance for identifying essential proteins no matter in which networks are.

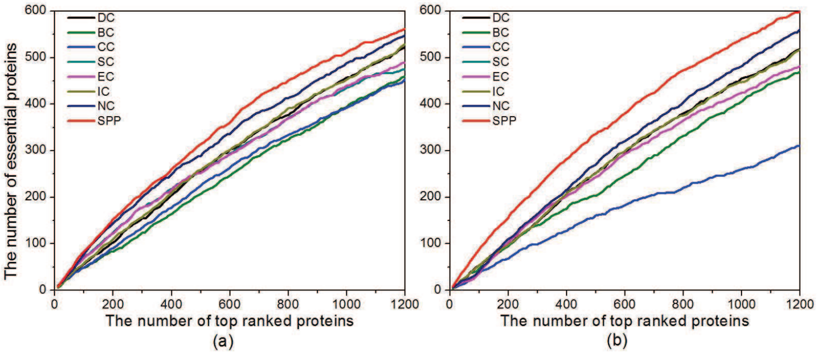


Figure 6: The performances of SPP and other methods on yeast PPI network: (a) YDIP; (b) YBioGRID

3.3. Difference analysis between SPP and seven representative methods

To further understand the reason why SPP has an outstanding performance, we perform a comparative analysis between SPP and seven network-based methods. Taking DC as example, we first selected the top 100 essential candidates of SPP and DC. There are 42 proteins identified by the two methods, and the number of proteins identified by SPP but not identified by DC is 58. Through comparing with the known essential protein data, only 36.2% proteins identified by DC are real essential proteins. By contrast, the real essential proteins identified by SPP account for 86.2%. The similar results of other methods are shown in Figure 7. The x axis represents each existing method, and the y axis is the percentage of essential proteins identified by these methods. We can see that SPP can have excellent performance over the seven network-based methods, which indicates the effective of integrating subcellular localization information and considering the highly interactions between essential proteins.

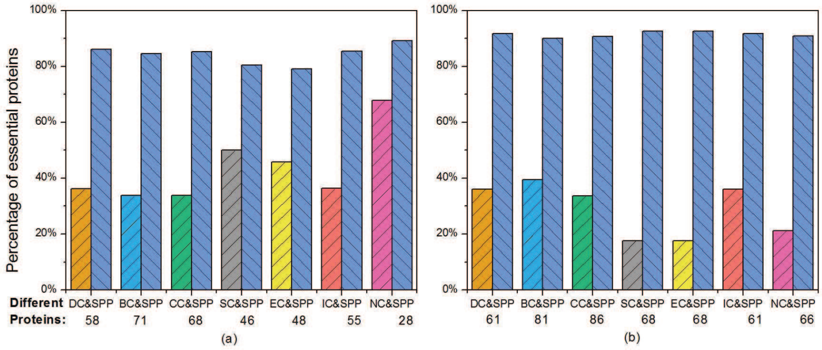


Figure 7: The percentage of essential proteins of the different proteins between SPP and other methods in the top 100 predictions from yeast PPI network: (a) YDIP; (b) YBioGRID

3.4. Validated by accuracy

Sensitivity (SN), Specificity (SP), Positive predictive value (PPV), Negative predictive value (NPV), F-measure (F) and Accuracy (ACC) are popular validation metrics for evaluating essential protein discovery methods. We also used these six metrics to validate the existing prediction methods and SPP. Let TP be the number of essential proteins and TN be the number of non-essential proteins which are correctly predicted, respectively, FN be the number of essential proteins which are missed, and FP be the number of proteins which are non-essential proteins but false predicted. Then these statistical measures are calculated as follows.

$$SN = TP/(TP + FN) \tag{3}$$

$$SP = TN/(FP + TN) \tag{4}$$

$$PPV = TP/(TP + FP) \tag{5}$$

$$NPV = TN/(TN + FN) \tag{6}$$

$$F = 2 \times SN \times PPV/(SN + PPV) \tag{7}$$

$$ACC = (TP + TN)/(TP + TN + FP + FN) \tag{8}$$

The comparison results of SN, SP, PPV, NPV, F-measure and ACC of SPP with seven other existing network-based methods are shown in Table 2. As shown in Table 2, no matter on YDIP or YBioGRID, we can get that the value of SN, SP, PPV, NPV, F and ACC of SPP are the highest among all methods, which indicates the effectiveness of this method for identifying essential proteins. 245

Table 2: **Comparison the result of SN, SP, PPV, NPV, F and ACC of SPP and other methods**

Dataset	Method	SN	SP	PPV	NPV	F	ACC
YDIP	DC	0.4398	0.8249	0.4398	0.8249	0.4398	0.7332
	BC	0.3885	0.8089	0.3885	0.8089	0.3885	0.7088
	CC	0.3858	0.8081	0.3858	0.8081	0.3858	0.7075
	SC	0.4142	0.8169	0.4142	0.8169	0.4142	0.721
	EC	0.4177	0.818	0.4177	0.818	0.4177	0.7227
	IC	0.4434	0.8261	0.4434	0.8261	0.4434	0.7349
	NC	0.4681	0.8338	0.4681	0.8338	0.4681	0.7467
	SPP	0.4823	0.8382	0.4823	0.8382	0.4823	0.7535
YBioGRID	DC	0.4329	0.846	0.4329	0.846	0.4329	0.7578
	BC	0.3928	0.8352	0.3928	0.8352	0.3928	0.7407
	CC	0.2602	0.7992	0.2602	0.7992	0.2602	0.6841
	SC	0.4012	0.8374	0.4012	0.8374	0.4012	0.7443
	EC	0.4012	0.8374	0.4012	0.8374	0.4012	0.7443
	IC	0.432	0.8458	0.432	0.8458	0.432	0.7575
	NC	0.4671	0.8553	0.4671	0.8553	0.4671	0.7724
	SPP	0.5004	0.8644	0.5004	0.8644	0.5004	0.7867

4. Conclusions

Identifying essential proteins has important significance in biomedical science. Up to now, there are many network-based computational methods proposed and applied widely. However, because of defects of high-throughput technologies, the PPI data are not comprehensive and have high noise, which makes a great challenge to predict more essential proteins. Based on the characteristics of proteins, the original network can be divided into several compartment sub-networks by integrating subcellular localization information, in which we can analyze the connection relations of protein nodes using common neighbors. As to the analysis above, we propose a new computational method to identify essential proteins, named SPP. Through comparing with seven existing network-based methods, i.e. DC, BC, CC, SC, EC, IC and NC, SPP not only can effectively improve the accuracy for identifying essential proteins, but also has a better performance both on YDIP and YBioGRID datasets.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants (No.61622213, No.61728211, No.61772552, and No.61420106009) and the Fundamental Research Funds for the Central Universities of Central South University (No.2017zzts504).

Appendix A. Supplementary data

The PPI data, subcellular localization information and list of essential proteins are available at *http://bioinformatics.csu.edu.cn/resources/softs/spp/*. The detailed information of sub-networks and experimental results also can be downloaded from the webpage.

References

References

[1] R. S. Kamath, A. G. Fraser, Y. Dong, G. Poulin, et al., Systematic functional analysis of the caenorhabditis elegans genome using rnai, Nature 421 (6920) (2003) 231.

[2] C. Pál, B. Papp, L. D. Hurst, Genomic function (communication arising): Rate of evolution and gene dispensability, Nature 421 (6922) (2003) 496–497.

[3] E. A. Winzeler, D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J. D. Boeke, H. Bussey, et al., Functional characterization of the s. cerevisiae genome by gene deletion and parallel analysis, science 285 (5429) (1999) 901–906.

[4] J. I. Glass, C. A. Hutchison, H. O. Smith, J. C. Venter, A systems biology tour de force for a near-minimal bacterium, Molecular systems biology 5 (1) (2009) 330.

[5] W. Lan, J. Wang, M. Li, W. Peng, F.-X. Wu, Computational approaches for prioritizing candidate disease genes based on ppi networks, Tsinghua Science and Technology 20 (5) (2015) 500–512.

[6] G. Giaever, A. M. Chu, L. Ni, C. Connelly, et al., Functional profiling of the saccharomyces cerevisiae genome, nature 418 (6896) (2002) 387.

[7] L. M. Cullen, G. M. Arndt, Genome-wide screening for gene function using rnai in mammalian cells, Immunology and cell biology 83 (3) (2005) 217.

[8] T. Roemer, B. Jiang, J. Davison, T. Ketela, K. Veillette, A. Breton, F. Tandia, A. Linteau, S. Sillaots, C. Marta, et al., Large-scale essential gene identification in candida albicans and applications to antifungal drug discovery, Molecular microbiology 50 (1) (2003) 167–181.

[9] S. Fields, O.-k. Song, A novel genetic system to detect protein–protein interactions, Nature 340 (6230) (1989) 245–246.

[10] Y. Ho, A. Gruhler, A. Heilbut, G. D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, et al., Systematic identification of protein complexes in saccharomyces cerevisiae by mass spectrometry, Nature 415 (6868) (2002) 180–183.

[11] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, et al., Global analysis of protein activities using proteome chips, science 293 (5537) (2001) 2101–2105.

[12] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Séraphin, A generic protein purification method for protein complex characterization and proteome exploration, Nature biotechnology 17 (10) (1999) 1030–1032.

[13] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S.-M. Kim, D. Eisenberg, Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions, Nucleic acids research 30 (1) (2002) 303–305.

[14] A. Chatr-Aryamontri, B.-J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, C. Stark, A. Breitkreutz, N. Kolas, L. O'donnell, et al., The biogrid interaction database: 2015 update, Nucleic acids research 43 (D1) (2014) D470–D478.

[15] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico, et al., Mint, the molecular interaction database: 2012 update, Nucleic acids research 40 (D1) (2011) D857–D861.

[16] G. D. Bader, D. Betel, C. W. Hogue, Bind: the biomolecular interaction network database, Nucleic acids research 31 (1) (2003) 248–250.

[17] E. Estrada, Virtual identification of essential proteins within the protein interaction network of yeast, Proteomics 6 (1) (2006) 35–40.

[18] W. Peng, J. Wang, W. Wang, Q. Liu, F.-X. Wu, Y. Pan, Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks, BMC systems biology 6 (1) (2012) 87.

[19] M. Li, R. Zheng, H. Zhang, J. Wang, Y. Pan, Effective identification of essential proteins based on priori knowledge, network topology and gene expressions, Methods 67 (3) (2014) 325–333.

[20] H. Jeong, Z. N. Oltvai, A.-L. Barabasi, Prediction of protein essentiality based on genomic data, ComPlexUs 1 (1) (2003) 19–28.

[21] C.-C. Lin, H.-F. Juan, J.-T. Hsiang, Y.-C. Hwang, H. Mori, H.-C. Huang, Essential core of protein- protein interaction network in escherichia coli, Journal of proteome research 8 (4) (2009) 1925–1931.

[22] M. W. Hahn, A. D. Kern, Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks, Molecular biology and evolution 22 (4) (2004) 803–806.

[23] H. Liang, W.-H. Li, Gene essentiality, gene duplicability and protein connectivity in human and mouse, Trends in Genetics 23 (8) (2007) 375–378.

[24] E. Estrada, J. A. Rodriguez-Velazquez, Subgraph centrality in complex networks, Physical Review E 71 (5) (2005) 056103.

[25] P. Bonacich, Power and centrality: A family of measures, American journal of sociology 92 (5) (1987) 1170–1182.

[26] C.-Y. Lin, C.-H. Chin, H.-H. Wu, S.-H. Chen, C.-W. Ho, M.-T. Ko, Hubba: hub objects analyzera framework of interactome hubs identification for network biology, Nucleic acids research 36 (suppl_2) (2008) W438–W443.

[27] M. Li, J. Wang, X. Chen, H. Wang, Y. Pan, A local average connectivity-based method for identifying essential proteins from the network level, Computational biology and chemistry 35 (3) (2011) 143–150.

[28] J. Wang, M. Li, H. Wang, Y. Pan, Identification of essential proteins based on edge clustering coefficient, IEEE/ACM Transactions on Computational Biology and Bioinformatics 9 (4) (2012) 1070–1080.

[29] M. Li, Y. Lu, J. Wang, F.-X. Wu, Y. Pan, A topology potential-based method for identifying essential proteins from ppi networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 12 (2) (2015) 372–383.

[30] M. P. Joy, A. Brock, D. E. Ingber, S. Huang, High-betweenness proteins in the yeast protein interaction network, BioMed Research International 2005 (2) (2005) 96–103.

[31] S. Wuchty, P. F. Stadler, Centers of complex networks, Journal of Theoretical Biology 223 (1) (2003) 45–53.

[32] K. Stephenson, M. Zelen, Rethinking centrality: Methods and examples, Social networks 11 (1) (1989) 1–37.

[33] N. Pržulj, D. A. Wigle, I. Jurisica, Functional topology in a network of protein interactions, Bioinformatics 20 (3) (2004) 340–348.

[34] Y. Tang, M. Li, J. Wang, Y. Pan, F.-X. Wu, Cytonca: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks, Biosystems 127 (2015) 67–72.

[35] M. Li, J. Yang, F.-X. Wu, Y. Pan, J. Wang, Dynetviewer: a cytoscape app for dynamic network construction, analysis and visualization, Bioinformatics 1 (2018) 3.

[36] M. Cheng, L. Liu, H. Wang, C. Du, W. Song, Essential proteins discovery from weighted protein–protein interaction networks, Journal of Bionanoscience 8 (4) (2014) 293–297.

[37] X. Tang, J. Wang, J. Zhong, Y. Pan, Predicting essential proteins based on weighted degree centrality, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 11 (2) (2014) 407–418.

[38] M. Li, J.-X. Wang, H. Wang, Y. Pan, Identification of essential proteins from weighted protein–protein interaction networks, Journal of bioinformatics and computational biology 11 (03) (2013) 1341002.

[39] X. Peng, J. Wang, J. Wang, F.-X. Wu, Y. Pan, Rechecking the centrality-lethality rule in the scope of protein subcellular localization interaction networks, PloS one 10 (6) (2015) e0130743.

[40] M. Li, P. Ni, X. Chen, J. Wang, F.-X. Wu, Y. Pan, Construction of refined protein interaction network for predicting essential proteins, IEEE/ACM transactions on computational biology and bioinformatics.

[41] M. Li, X. Meng, R. Zheng, F.-X. Wu, Y. Li, Y. Pan, J. Wang, Identification of protein complexes by using a spatial and temporal active protein interaction network, IEEE/ACM transactions on computational biology and bioinformatics.

[42] M. Li, Z. Niu, X. Chen, P. Zhong, F.-X. Wu, Y. Pan, A reliable neighbor-based method for identifying essential proteins by integrating gene expressions, orthology, and subcellular localization information, Tsinghua Science and Technology 21 (6) (2016) 668–677.

[43] W. Zhang, J. Xu, Y. Li, X. Zou, Detecting essential proteins based on network topology, gene expression data and gene ontology information, IEEE/ACM transactions on computational biology and bioinformatics.

[44] 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 (1) (2009) 290.

[45] G. Li, M. Li, J. Wang, J. Wu, F.-X. Wu, Y. Pan, Predicting essential proteins based on subcellular localization, orthology and ppi networks, BMC bioinformatics 17 (8) (2016) 279.

[46] Q. Xiao, J. Wang, X. Peng, F.-X. Wu, Y. Pan, Identifying essential proteins from active ppi networks constructed with dynamic gene expression, BMC genomics 16 (3) (2015) S1.

[47] J. Wang, X. Peng, M. Li, Y. Pan, Construction and application of dynamic protein interaction network based on time course gene expression data, Proteomics 13 (2) (2013) 301–312.

[48] M. Li, Y. Lu, Z. Niu, F.-X. Wu, United complex centrality for identification of essential proteins from ppi networks, IEEE/ACM transactions on computational biology and bioinformatics 14 (2) (2017) 370–380.

[49] J. Luo, Y. Qi, Identification of essential proteins based on a new combination of local interaction density and protein complexes, PloS one 10 (6) (2015) e0131418.

[50] A.-C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, et al., Functional organization of the yeast proteome by systematic analysis of protein complexes, Nature 415 (6868) (2002) 141–147.

[51] X. Peng, J. Wang, J. Huan, F.-X. Wu, Double-layer clustering method to predict protein complexes based on power-law distribution and protein sublocalization, Journal of theoretical biology 395 (2016) 186–193.

[52] A. Kumar, S. Agarwal, J. A. Heyman, S. Matson, M. Heidtman, S. Piccirillo, L. Umansky, A. Drawid, R. Jansen, Y. Liu, et al., Subcellular localization of the yeast proteome, Genes & development 16 (6) (2002) 707–719.

[53] M. S. Scott, S. J. Calafell, D. Y. Thomas, M. T. Hallett, Refining protein subcellular localization, PLoS computational biology 1 (6) (2005) e66.

[54] W. Peng, M. Li, L. Chen, L. Wang, Predicting protein functions by using unbalanced random walk algorithm on three biological networks, IEEE/ACM transactions on computational biology and bioinformatics 14 (2) (2017) 360–369.

[55] W.-K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, et al., Global analysis of protein localization in budding yeast, Nature 425 (6959) (2003) 686.

[56] X. Peng, J. Wang, W. Peng, F.-X. Wu, Y. Pan, Protein–protein interactions: detection, reliability assessment and applications, Briefings in bioinformatics.

[57] R. Zhang, H.-Y. Ou, C.-T. Zhang, Deg: a database of essential genes, Nucleic acids research 32 (suppl_1) (2004) D271–D272.

[58] 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 research 30 (1) (2002) 31–34.

[59] L. Issel-Tarver, K. R. Christie, K. Dolinski, R. Andrada, R. Balakrishnan, C. A. Ball, G. Binkley, S. Dong, S. S. Dwight, D. G. Fisk, et al., Saccharomyces genome database, Methods in enzymology 350 (2002) 329–346.

[60] Saccharomyces genome deletion project, http://www-sequence.stanford.edu/group/yeast_deletion_project/.

[61] J. X. Binder, S. Pletscher-Frankild, K. Tsafou, C. Stolte, S. I. ODonoghue, R. Schneider, L. J. Jensen, Compartments: unification and visualization of protein subcellular localization evidence, Database 2014 (2014) bau012.

[62] S. Briesemeister, J. Rahnenführer, O. Kohlbacher, Going from where to whyinterpretable prediction of protein subcellular localization, Bioinformatics 26 (9) (2010) 1232–1238.

[63] S. Park, J.-S. Yang, Y.-E. Shin, J. Park, S. K. Jang, S. Kim, Protein localization as a principal feature of the etiology and comorbidity of genetic diseases, Molecular systems biology 7 (1) (2011) 494.

[64] M. Dreger, Subcellular proteomics, Mass spectrometry reviews 22 (1) (2003) 27–56.

[65] J. B. Pereira-Leal, B. Audit, J. M. Peregrin-Alvarez, C. A. Ouzounis, An exponential core in the heart of the yeast protein interaction network, Molecular biology and evolution 22 (3) (2004) 421–425.

[66] G. Butland, J. M. Peregrín-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, et al., Interaction network containing conserved and essential protein complexes in escherichia coli, Nature 433 (7025) (2005) 531–537.