IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 12, 2020, accepted July 26, 2020, date of publication July 29, 2020, date of current version August 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3012682

A Novel Network-Based Computational Model for Prediction of Essential Proteins

XIANYOU ZHU[1], YANG LIU[2,3], TINGRUI PEI[3], ZHIPING CHEN[2], XUEYONG LI[2], AND WANG LEI[2,3]

[1]Hunan Provincial Key Laboratory of Intelligent Information Processing and Application, Hengyang Normal University, Hengyang 421002, China
[2]College of Computer Engineering and Applied Mathematics, Changsha University, Changsha 410022, China
[3]Key Laboratory of Hunan Province for Internet of Things and Information Security, Xiangtan University, Xiangtan 411105, China

Corresponding authors: Yang Liu (y1006480772@163.com) and Wang Lei (wanglei@xtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61873221 and Grant 61672447; in part by the Natural Science Foundation of Hunan Province under Grant 2019JJ70010; in part by the Science and Technology Plan Project of Hunan Province under Grant 2016TP1020 and Grant 2019TP1011; and in part by the Double First-Class University Project of Hunan Province under Grant Xi'an Jiaotong [2018]469.

ABSTRACT Identification of essential proteins is important for understanding cell survival and development, because even if only one of these proteins is missing, organisms cannot survive or develop. Since traditional methods for identifying essential proteins based on biological experiments are costly and inefficient, more and more computational models are proposed for predicting essential proteins in recent years. In this paper, a novel computational model called BSPM is proposed, in which, an original PPI network will be built based on known protein-protein associations first, and then topology information of the original PPI network will be adopted to measure the similarities between proteins based on the SimRank algorithm. Thereafter, a weighted PPI network can be obtained based on the similarities between proteins and the original PPI network. Finally, based on the weighted PPI network, the PageRank algorithm will be used to infer potential essential proteins. Moreover, in order to evaluate the performance of BSPM, we have compared the performance of BSPM with 14 classical prediction models in the field based on two different databases, and experimental results show that BSPM can achieve prediction accuracies of 92%, 81% and 76% out of the top 100, 200 and 300 candidate proteins separately, which not only are significantly better than those 14 competitive classical prediction models, but also means that BSPM can be used as an effective model for identifying essential proteins in the future.

INDEX TERMS Essential protein, PPI, SimRank algorithm, PageRank algorithm.

I. INTRODUCTION

More and more evidences show that proteins are involved in almost all life activities, but different proteins have different functions and importance in life activities. As an important proteome, essential proteins play a vital role in the development and survival of organisms. In theory, identification of essential proteins can not only provide insights into the minimum requirements for cell survival and development, but also play important roles in the emerging synthetic biology science, which aims to create cells with the smallest genome [1]. From a practical point of view, essential proteins have become drug targets for new antibiotics due to their indispensability for bacterial cell survival [2]. In biology, there are many experimental methods that can predict and discover essential proteins, such as single gene knockout [3], RNA interference [4] and conditional knockout [5]. However, these experiments are expensive and inefficient. In addition, they are limited to a few species. Therefore, high-precision calculation methods have become a very important choice for identifying essential proteins.

Existing essential protein prediction models can be roughly divided into two categories. The first model predicts key proteins based on the topological characteristics of the PPI network. For example, the centrality-lethality rule proposed by Jeong *et al.* Including degree centrality (DC) [6], intermediate degree centrality (BC) [7], proximity centrality (CC) [8], subgraph centrality (SC) [9], feature vector centrality (EC) [10], Information Center (IC) [11]. BC is a global metric used to calculate the proportion of the shortest path

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

through a given node. CC is also a global metric, which evaluates the closeness of all remaining proteins in the network where a given node interacts with a given protein. The SC is responsible for the participation of nodes in all subgraphs of the network. EC simulates a mechanism in which each node affects all neighbors in the network, and IC describes how information flows through many different paths. These methods rank proteins according to their central position in the PPI network. Then, use the ranking scores of these proteins to determine whether the protein is essential. The advantage of these methods is that they can directly identify essential proteins without having to train a classifier based on a known set of essential proteins.

The second type of model infers key proteins by combining PPI networks with some biological information, including subcellular localization, evolutionary conservation, and gene expression. Yu *et al.* [12] studied the importance of bottlenecks in PIN and studied the correlation between PIN and the nature of genes by constructing the shortest path tree starting from each node in the network. Wang *et al.* [13] proposed a neighborhood-based method NC. This method identifies essential proteins based on the number of neighbors the protein has and the edge clustering coefficients that connect the interaction between the protein and its neighbors. Lei *et al.* [14] proposed a computational model RSG, which identifies key proteins based on a novel weighted PPI network constructed based on information from RNA-Seq, subcellular localization, and GO annotation data sets. Based on the assumption that key proteins tend to form dense clusters, Shabnam and Izudheen [15] constructed a prediction model by integrating gene expression profiles and domain information. Zhao *et al.* [16] constructed a weighted network based on gene expression data and topological information of the weighted network, and designed a calculation method named POEM based on this network to predict key proteins based on overlapping modules. In addition, Zhao *et al.* [17] proposed a new method called PeC by integrating network topology and gene expression, which increases the predictability of essential proteins compared to the concentration measurement based solely on network topology. Wang *et al.* [18] developed a computational model that predicts essential proteins based on PPI network topological properties and biological information (including subcellular localization data and orthologous data).

In this paper, a predictive model called BSPM is proposed, in which, original PPI networks are first constructed based on known associations between 1855 proteins downloaded from the Gavin database and known association between 5093 proteins downloaded from the DIP database separately. And then, the SimRank algorithm is adopted to calculate similarities between protein nodes in these original PPI networks based on the assumption that if two nodes are similar, then the nodes relating to them shall be similar as well. Thereafter, based on the newly obtained similarities between proteins, a weighted PPI network is constructed, in which, the problem caused by the sparseness of known associations between proteins can be solved effectively. Finally, based on the weighted PPI network, the PageRank algorithm is introduced to infer potential essential proteins through iterative propagation. In addition, in order to evaluate the performance of BSPM, based on the Gavin database and the DIP database, we have compared BSPM with 14 classical methods, including DC [6], BC [7], CC [8], SC [9], EC [10], IC [11], POEM [16], RWHN [17], NC [18], PEC [19], CoEWC [20], ION [21], LAC [22] and NPRI [23] respectively, and experimental results show that BSPM can achieve prediction accuracy of 92%, 81% and 76% in the top 100, top 200 and top 300 candidate essential proteins separately, which are roughly better than all those 14 competitive state-of-the-art models. Moreover, the simulation results of parameter analysis based on the Gavin dataset show that the model performance of BSPM is less affected by the parameter values, which indicates that BSPM has good stability. Hence, it is reasonable to draw a conclusion that BSPM can be used as an effective mean for identifying essential proteins in the future.

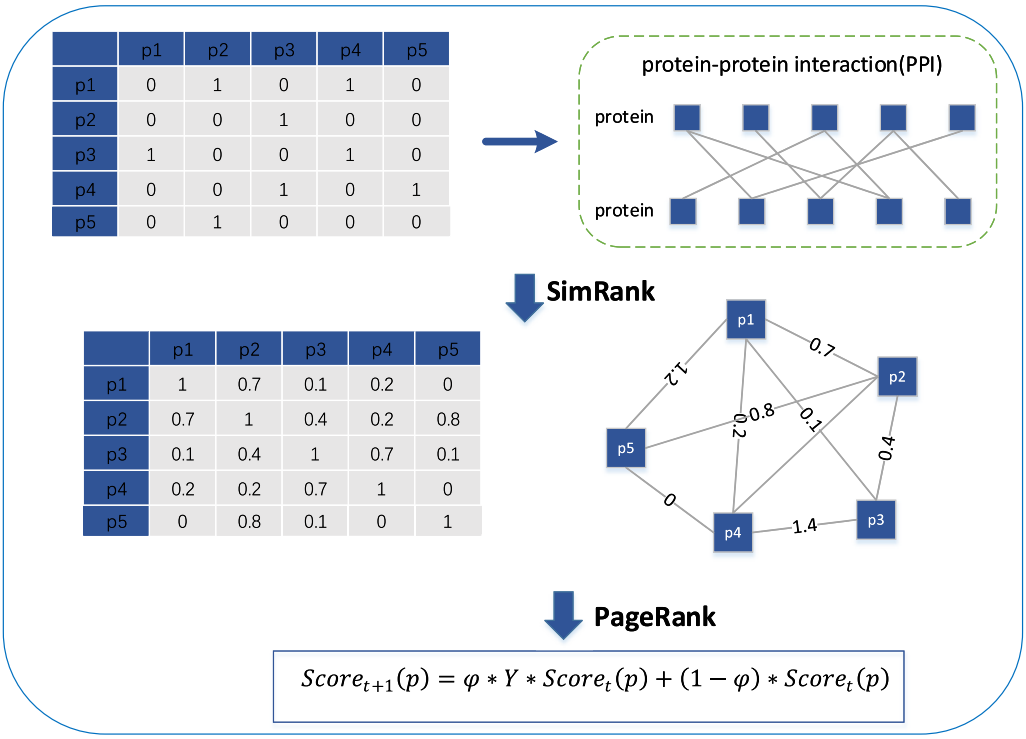
II. METHOD

A. CONSTRUCTION OF ORIGINAL PPI NETWORKS

In this section, we downloaded a dataset of known protein-protein associations from the Gavin *et al.* [25] database first. After screening, we finally obtained 7,669 known protein-protein associations between 617 essential proteins and 1855 proteins. Next, we downloaded another dataset of known protein-protein associations from the DIP [26] database as well, and after screening, we finally obtained 24,743 known protein-protein associations between 1167 essential proteins and 5093 proteins. Thereafter, we adopted an adjacency matrix $A = \{a_{ij}\}$ to describe the dataset of known associations between proteins downloaded from any given database, in which, for any two proteins P_i and P_j, if there is a known association between them in the newly downloaded dataset of known associations between proteins, then there is $a_{ij} = 1$, otherwise there is $a_{ij} = 0$. Moreover, let N_p represent the number of proteins in the newly downloaded dataset, it is obvious that we can construct a protein-protein interaction (PPI) network $G(V, E)$ conveniently, where $V = \{v_1, \cdots, v_{N_p}\}$ represents the set of different proteins, and $E \in V \times V$ denotes the set of known protein-protein interactions.

B. SIMILARITIES OF PROTEINS BASED ON COMMON NEIGHBORS

In $G(V, E)$, $\forall p \in V$, let $\theta(p)$ denote the set of neighboring nodes of p, and $d(p)$ represent the degree of p, then for any two nodes p_i and p_j in V, if there are common neighboring nodes between them, obviously, it is reasonable to assume that there is potential relationship between them. Hence, we can define the potential similarity between p_i and



	p1	p2	p3	p4	p5
p1	0	1	0	1	0
p2	0	0	1	0	0
p3	1	0	0	1	0
p4	0	0	1	0	1
p5	0	1	0	0	0

	p1	p2	p3	p4	p5
p1	1	0.7	0.1	0.2	0
p2	0.7	1	0.4	0.2	0.8
p3	0.1	0.4	1	0.7	0.1
p4	0.2	0.2	0.7	1	0
p5	0	0.8	0.1	0	1

FIGURE 1. The flowchart of BSPM.

p_j as follows:

$$Sim\,(i,j) = \begin{cases} 1: & If\ i=j \\ 0: & Else\ if\ \theta\,(p_i)=\emptyset or \theta\,(p_j)=\emptyset \\ \dfrac{1}{d(p_i)\times d(p_j)} & \\ \sum_{r=1}^{N_p}\left(a_{ir}\times a_{jr}\right): & Otherwise \end{cases} \tag{1}$$

C. SIMILARITIES BETWEEN PROTEINS BASED ON THE SimRank

The similarity between two nodes p_i and p_j defined by the SimRank model [27] is based on the recursive idea that if those nodes neighboring to both p_i and p_j are similar, then these two nodes p_i and p_j can be considered to be similar to each other as well. Based on such concept of the SimRank model, it is obvious that for any two nodes p_i and p_j in V, if there are no common neighboring nodes between them, but they are directly connected to a pair of similar nodes, then we can assume that there is potential similarity between them as well. Here, the potential similarity between these two nodes p_i and p_j can be defined as follows:

$$S_k\,(i,j) = \begin{cases} 1: & If\ i=j \\ 0: & Else\ if\ \theta\,(p_i)=\emptyset or \theta\,(p_j)=\emptyset \\ \dfrac{1}{d(p_i)\times d(p_j)} & \\ \sum_{m=1}^{N_p}\sum_{n=1}^{N_p} a_{mi}\times a_{nj} & \\ \times S_{k-1}\,(m,n): & Otherwise \end{cases} \tag{2}$$

Here, $S_0(i,j) = Sim(i,j)$ and k is a parameter indicating the number of iterations. Based on above formula (2), it is easy to see that we can further integrate the newly obtained similarity matrix S and the previously obtained original adjacency matrix A in the following way:

$$Y = \begin{bmatrix} S & A \\ A^T & S \end{bmatrix} \tag{3}$$

Thereafter, a new integrated matrix Y can be obtained, which can not only be used for further essential protein prediction, but also guarantee that BSPM is able to be applied for identifying potential essential proteins without known associations. Moreover, based on the integrated matrix Y, it is obvious that a novel weighted PPI network can be constructed easily as well.

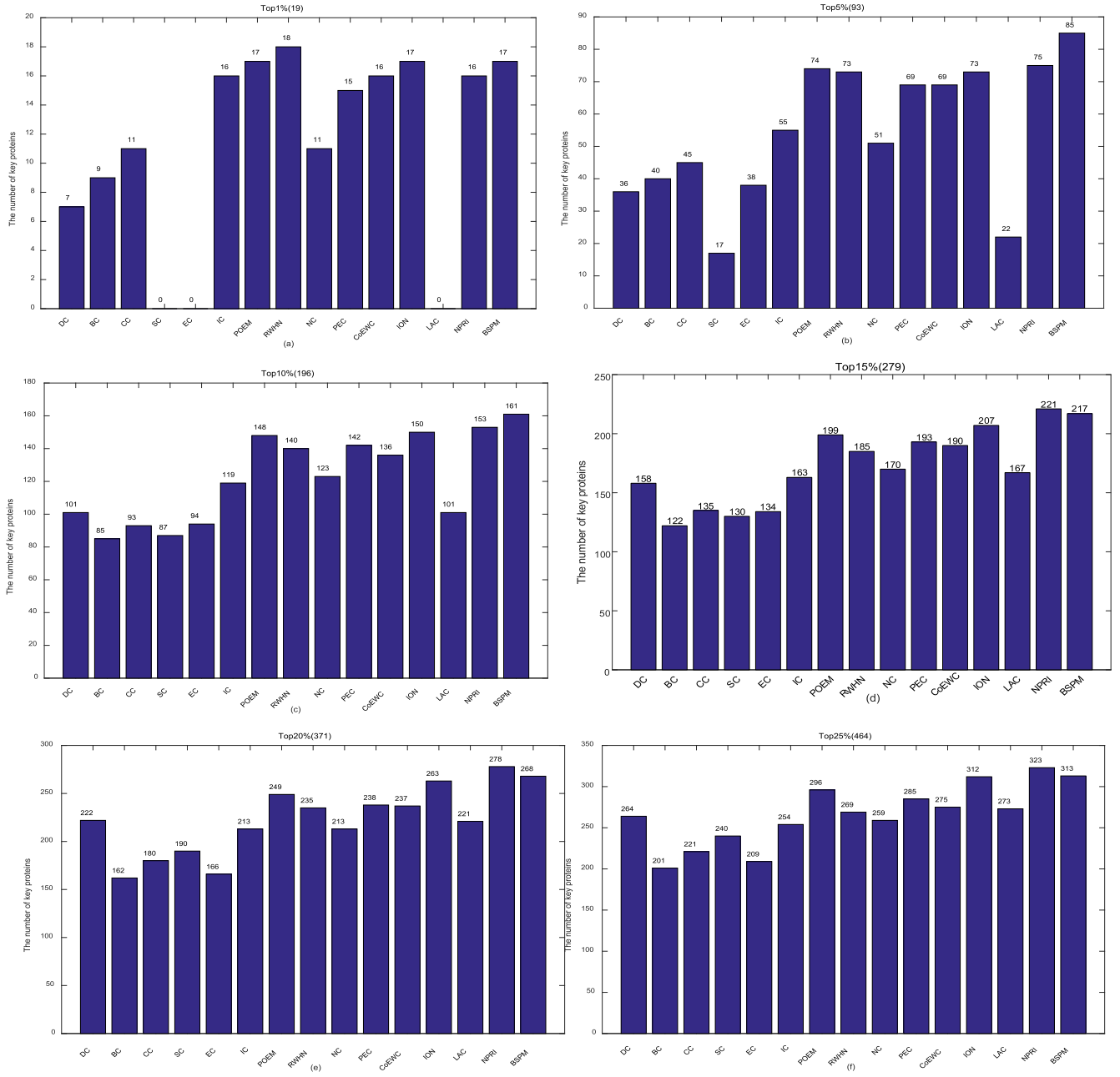


FIGURE 2. Comparison results obtained based on the Gavin database. (a) The number of essential proteins in the top 1% ranked proteins. (b) The number of essential proteins in the top 5% ranked proteins. (c) The number of essential proteins in the top 10% ranked proteins. (d) The number of essential proteins in the top 15% ranked proteins. (e) The number of essential proteins in the top 20% ranked proteins. (f) The number of essential proteins in the top 25% ranked proteins.

D. CALCULATION OF INITIAL SCORES FOR PROTEINS

In order to assign initial scores for protein nodes in the newly constructed weighted PPI network, we further downloaded the orthology information from the InParanoid database [28] in this section. For each protein node $p_i \in V$, let $ort(p_i)$ denote the conservative score downloaded from the InParanoid database, then we can define the characteristic value of its orthogonal information as follows:

$$pro_{ort(p_i)} = \frac{ort(p_i)}{max_{p_j \in V}(ort(p_j))} \tag{4}$$

E. CALCULATION OF RANKING SCORES

Based on the PageRank algorithm, in this section, a novel computational algorithm called BSPM is proposed for predicting potential essential proteins. The flowchart of BSPM is shown in the following Figure 1. In BSPM, an original PPI network is built first based on newly downloaded known associations between proteins. And then, based on the SimRank model, the original PPI network will be transformed to a weighted PPI network. Next, each node in the weighted PPI network will calculate an initial score value according to

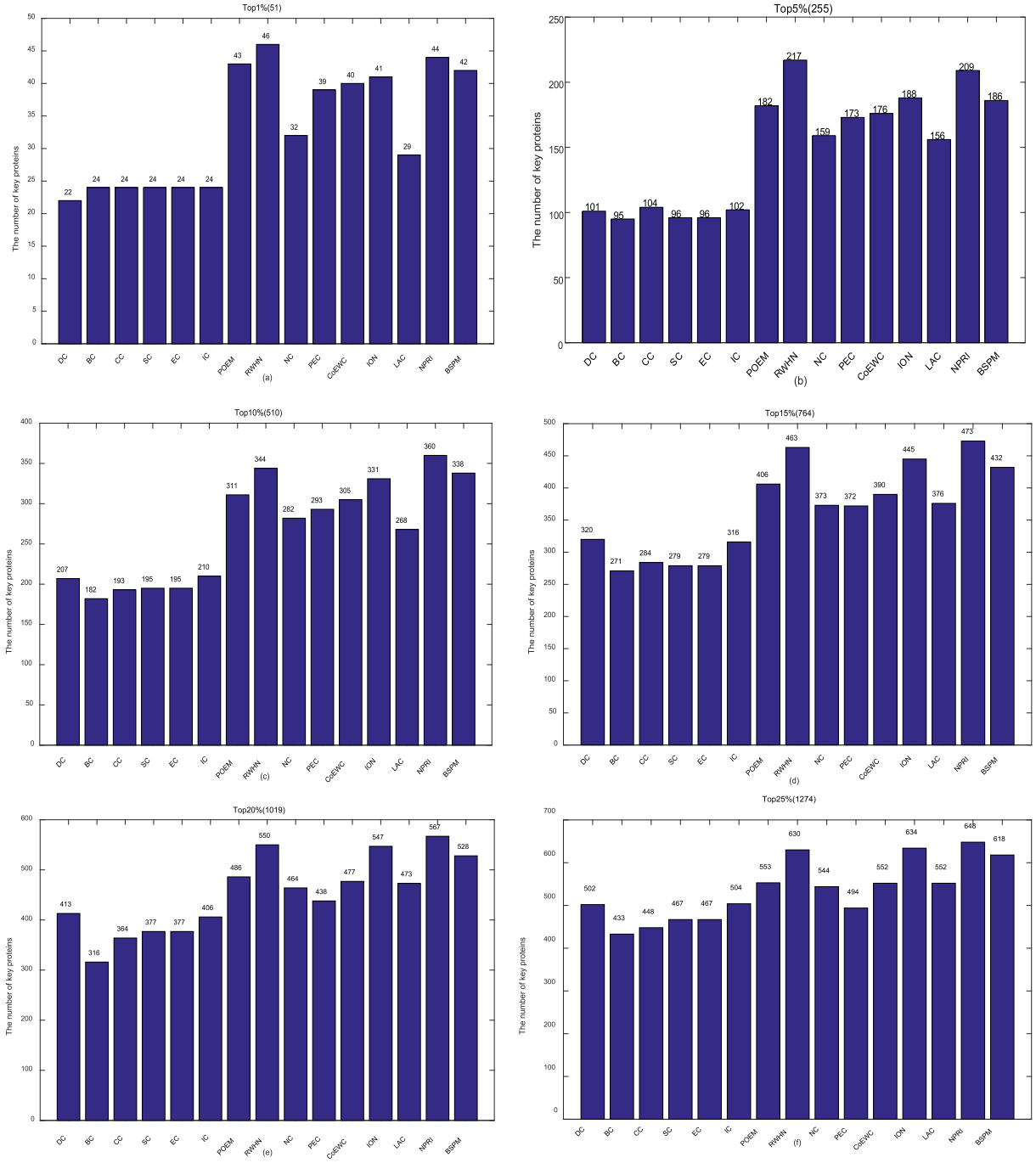


FIGURE 3. Comparison results obtained based on the DIP database. (a) The number of essential proteins in the top 1% ranked proteins. (b) The number of essential proteins in the top 5% ranked proteins. (c) The number of essential proteins in the top 10% ranked proteins. (d) The number of essential proteins in the top 15% ranked proteins. (e) The number of essential proteins in the top 20% ranked proteins. (f) The number of essential proteins in the top 25% ranked proteins.

above formula (4), based on which, each node will obtain a final score value iteratively based on the PageRank algorithm as follows:

$$Score_{t+1}(p) = \varphi * Y * Score_t(p) + (1-\varphi) * Score_t(p) \quad (5)$$

Here, $\forall p \in V$, $Score_0(p) = pro_{ort(p_i)}$, and $Score_t(p)$ denotes score obtained by the node p at its t-th round of iteration.

III. RESULTS AND ANALYSIS

A. COMPARISON WITH COMPETITIVE PREDICTION METHODS

To evaluate the performance of BSPM, we compared it with 14 competitive prediction methods, such as DC [6], BC [7], CC [8], SC [9], EC [10], IC [11], POEM [16], RWHN [17], NC [18], PEC [19], CoEWC [20], ION [21], LAC [22], NPRI [23] respectively. Here, in POEM, the original

TABLE 1. Comparison between BSPM and competitive methods based on the top 1% predicted proteins.

	DC	BC	CC	SC	EC	IC	POEM	RWHN	NC	PEC	CoEWC	ION	LAC	NPRI	BSPM
Gavin	0.368	0.474	0.579	0	0	0.842	0.895	0.947	0.579	0.789	0.842	0.895	0	0.842	0.895
DIP	0.431	0.471	0.471	0.471	0.471	0.471	0.843	0.902	0.627	0.784	0.804	0.804	0.569	0.863	0.824

TABLE 2. The influence of φ on the prediction performance of BSPM based on the Gavin database.

φ	0.1	0.2	0.3	0.4	0.5
1%（19）	0.89	0.89	0.89	0.84	0.78
5%（93）	0.91	0.88	0.82	0.80	0.80
10%（196）	0.82	0.81	0.81	0.81	0.80
15%（279）	0.77	0.77	0.78	0.77	0.77
20%（371）	0.72	0.71	0.71	0.71	0.70
25%（464）	0.67	0.66	0.66	0.66	0.66

protein interaction group network is divided into many overlapping basic modules first, and then essential proteins are predicted based on these overlapping basic modules. However, in RWHN, a transition probability matrix is built first, and then, the PageRank algorithm is adopted to predict the necessary proteins based on the original PPI network and known protein-domain associations. Different from these two kinds of prediction models mentioned above, NC is a new method for measuring the centrality of important protein recognition based on edge clustering coefficients, in which, both the centrality of nodes and their relationship with neighbors are considered. Similar to NC, PEC is also a centrality based model. However, different from NC, PEC is designed on the basis of both protein-protein interactions and gene expression data.

In CoEWC (Co-Expression Weighted by clustering Coefficients), essential proteins are predicted based on the topology of PPI networks and co-expression of interacting proteins. ION identifies essential proteins by integrating the orthology information with the PPI networks. LAC determines the necessity of a protein by evaluating the relationship between the protein and its neighbors. NPRI recognizes essential proteins based on protein-domain networks and domain-domain networks and protein-protein association networks.

During simulation, the performance of each method is judged based on the number of truly essential proteins identified by the method. Experiments are executed on the basis of the Gavin database and the DIP database separately, and experimental results are shown in Figure 2 and Figure 3 respectively. During simulation, proteins are ranked in descending order based on their ranking scores calculated by BSPM and 14 competitive methods respectively. And then, the top 1%, top 5%, top 10%, top 15%, top 20% and top 25% ranked proteins will be chosen as candidate essential proteins. Thereafter, through comparing with known essential proteins, the number of true essential proteins detected by each method will be used as the judgment criteria of prediction ability. From observing Figure 2 and Figure 3, it is easy to see from the perspective of the Gavin database that the performance of BSPM is slightly lower than that of RWHN in the prediction of the top 1% essential proteins, and slightly lower than that of NPRI in the prediction of the top 15%, top 20% and top 25% essential proteins, but is higher than all these 14 competitive methods in the prediction of the top 5% and top 10% essential proteins. In addition, it can be seen from the perspective of the DIP database that the performance of BSPM is only lower than that of NPRI, RWHN and POEM while comparing with these 14 state-of-the-art models. Through analysis, the reason may be that the domain-related information is adopted in NPRI, RWHN and POEM. In addition, we have compared the prediction performance of BSPM with these 14 methods based on the top 1% ranked proteins, and the comparative results are shown in the Table 1.

B. EFFECTS OF PARAMETER α

In BSPM, we defined a parameter φ with a value between 0 and 1, which is used to adjust the proportion allocated during the iteration. Table 2 shows the prediction results based on the Gavin database when assigning different values to φ. Obviously, the prediction performance of BSPM varies with different φ values. In general, as the value of φ increases, the prediction performance of BSPM gradually decreases. It is easy to see that when the value of φ is 0.1, BSPM can achieve the best prediction performance.

IV. CONCLUSION AND DISCUSSION

Essential proteins are important for the survival and development of an organism, because even if only one of these proteins is missing, the organism cannot grow normally.

Traditional methods of identifying essential proteins through biological experiments are expensive and inefficient. Therefore, more and more computational models are proposed for predicting essential proteins. This paper proposes a prediction model called BSPM, in which, a protein-protein association network based on known protein-protein associations is built first. And then, based on the assumption that if two nodes are similar to each other, then these nodes related to them will be similar to each other as well, a SimRank algorithm is adopted to obtain protein-protein similarity matrix, which can solve the problem of sparse similarity matrix. Next, a weighted heterogeneous network is constructed based on the newly obtained similarity matrix and protein-protein association network. Finally, the PageRank algorithm is used to infer potential essential proteins. In addition, in order to evaluate the performance of BSPM, we have compared its performance with 14 classic prediction models based on the Gavin database and the DIP database separately. And experimental results show that BSPM can be used as a powerful tool for predicting essential proteins.

The reason why BSPM achieves better performance is due to the following points: First, considering the sparse of known protein-protein association, which leads to the sparseness of the similarity matrix calculated based on common neighbors, we use the PageRank algorithm to obtain protein-protein similarity matrix through iterative propagation. Second, we use the orthology information of the InParanoid database as the initial information of each node. Of course, there are still some limitations in BSPM that need to be improved in the future. For example, domain-related information is not used in the model, which makes the model performance slightly lower than the recently proposed classic model. Moreover, the main concept in our method might be used to predict essential microRNA through using known miRNA-disease associations [29]–[33] etc.

REFERENCES

[1] J. I. Glass, C. A. Hutchison, H. O. Smith, and J. C. Venter, "A systems biology tour de force for a near–minimal bacterium," *Mol. Syst. Biol.*, vol. 5, no. 1, p. 330, Jan. 2009.

[2] C. Chao-Xi, L. Jun, and C. Zong-Xi, "Targeting virulence: A new paradigm for antimicrobial therapy," *China Animal Husbandry Vet. Med.*, vol. 3, no. 9, p. 541, 2011.

[3] G. Giaever, "Functional profiling of the Saccharomyces cerevisiae genome," *Nature*, vol. 418, pp. 387–391, Dec. 2002.

[4] L. M. Cullen and G. M. Arndt, "Genome–wide screening for gene function using RNAi in mammalian cells," *Immunol. Cell Biol.*, vol. 83, no. 3, pp. 217–223, Jun. 2005.

[5] T. Roemer, B. Jiang, J. Davison, T. Ketela, K. Veillette, A. Breton, F. Tandia, A. Linteau, S. Sillaots, C. Marta, N. Martel, S. Veronneau, S. Lemieux, S. Kauffman, J. Becker, R. Storms, C. Boone, and H. Bussey, "Large-scale essential gene identification in candida albicans and applications to antifungal drug discovery," *Mol. Microbiol.*, vol. 50, no. 1, pp. 167–181, Aug. 2003.

[6] M. W. Hahn and A. D. Kern, "Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks," *Molecular Biol. Evol.*, vol. 22, no. 4, pp. 803–806, Apr. 2005.

[7] M. P. Joy, A. Brock, E. Donald Ingber, and E. al, "High-Betweenness Proteins in the Yeast Protein Interaction Network," *J. Biomed. Biotechnol.*, vol. 2005, no. 2, p. 96, 2014.

[8] S. Wuchty and P. F. Stadler, "Centers of complex networks," *J. Theor. Biol.*, vol. 223, no. 1, pp. 45–53, Jul. 2003.

[9] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 5, May 2005, Art. no. 056103.

[10] C. M. Vineyard, S. J. Verzi, and M. K. Bernard, "A multi-modal network architecture for knowledge discovery," *Secur. Inform.*, vol. 10, no. 2, p. 20, 2012.

[11] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social Netw.*, vol. 11, no. 1, pp. 1–37, Mar. 1989.

[12] H. Yu, "The importance of bottlenecks in protein networks:Correlation with gene essentiality and expression dynamics," *PLoS Comput. Biol.*, vol. 3, no. 4, 2007, Art. no. e59.

[13] J. Wang, "Dentification of essential proteins based on edge clustering coefficient," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 4, pp. 1070—1080, Apr. 2012.

[14] X. Lei, "Predicting essential proteins based on RNA-Seq, subcellular localization and GO annotation datasets," *Knowl. Based Syst.*, vol. 151, pp. 136–148, Jul. 2018.

[15] C. B. F. Shabnam and S. Izudheen, "UDoGeC:Essential protein prediction using domain and gene expression profiles," *Procedia Comput. Sci.*, vol. 93, pp. 1003–1009, 2016.

[16] B. Zhao, J. Wang, M. Li, F.-X. Wu, and Y. Pan, "Prediction of essential proteins based on overlapping essential modules," *IEEE Trans. Nanobiosci.*, vol. 13, no. 4, pp. 415–424, Dec. 2014.

[17] B. Zhao, Y. Zhao, X. Zhang, Z. Zhang, F. Zhang, and L. Wang, "An iteration method for identifying yeast essential proteins from heterogeneous network," *BMC Bioinf.*, vol. 20, no. 1, Dec. 2019.

[18] J. Wang, M. Li, H. Wang, and Y. Pan, "Identification of essential proteins based on edge clustering coefficient," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 4, pp. 1070–1080, Jul. 2012.

[19] M. Li, H. Zhang, J.-X. Wang, and Y. Pan, "A new essential protein discovery method based on the integration of protein-protein interaction and gene expression data," *BMC Syst. Biol.*, vol. 6, no. 1, p. 15, 2012.

[20] X. Zhang, J. Xu, and W.-X. Xiao, "A new method for the discovery of essential proteins," *PLoS ONE*, vol. 8, no. 3, Mar. 2013, Art. no. e58763.

[21] W. Peng, J. Wang, W. Wang, Q. Liu, F.-X. Wu, and Y. Pan, "Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks," *BMC Syst. Biol.*, vol. 6, no. 1, pp. 1–17, 2012.

[22] M. Li, J. Wang, X. Chen, H. Wang, and Y. Pan, "A local average connectivity-based method for identifying essential proteins from the network level," *Comput. Biol. Chem.*, vol. 35, no. 3, pp. 143–150, Jun. 2011.

[23] Z. Chen, Z. Meng, C. Liu, X. Wang, L. Kuang, T. Pei, and L. Wang, "A novel model for predicting essential proteins based on heterogeneous protein-domain network," *IEEE Access*, vol. 8, pp. 8946–8958, 2020.

[24] C. Qin, Y. Sun, and Y. Dong, "A new computational strategy for identifying essential proteins based on network topological properties and biological information," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0182031.

[25] A.-C. Gavin, "Proteome survey reveals modularity of the yeast cell machinery," *Nature*, vol. 440, no. 7084, pp. 631–636, Mar. 2006.

[26] I. Xenarios, "DIP, the database of interacting proteins: A research tool for studying cellular networks of protein interactions," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 303–305, Jan. 2002.

[27] L. Lu and T. Zhou, "Link prediction in complex networks: A survey," *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, 2010.

[28] T. Y. R. Juvik, "SGD : Saccharomyces Genome Database," *Nucleic Acids Res. Sci.*, vol. 71, no. 1, p. 9, 1998.

[29] F. Song, C. Cui, L. Gao, and Q. Cui, "MIES: Predicting the essentiality of miRNAs with machine learning and sequence features," *Bioinformatics*, vol. 35, no. 6, pp. 1053–1054, 2019.

[30] X. Chen, J. Yin, J. Qu, and L. Huang, "MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction," *PLoS Comput Biol.*, vol. 14, no. 8, 2018, Art. no. e1006418.

[31] X. Chen and J. Yin, "Ensemble of decision tree reveals potential miRNA-disease associations," *PLoS Comput. Biol.*, vol. 15, no. 7, 2019, Art. no. e1007209.

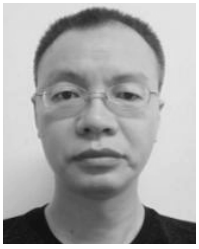
[32] X. Chen and L. Wang, "Predicting miRNA-disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, 2018.

[33] X. Chen, D. Xie, Q. Zhao, and Y. ZH, "MicroRNAs and complex diseases: From experimental results to computational models," *Brief Bioinform.*, vol. 20, no. 2, pp. 515–539, 2019.

XIANYOU ZHU received the B.S., M.S., and Ph.D. degrees in computer science and technology in 2000, 2005, and 2019, respectively. He is currently an Associate Professor with Hengyang Normal University. His current research interest includes bioinformatics.

YANG LIU is currently pursuing the master's degree in computer science and technology with the College of Information and Engineering, Xiangtan University. Her current research interest includes bioinformatics.

TINGRUI PEI received the B.S. and M.S. degrees from Xiangtan University, in 1992, and the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, in 2004. From 2006 to 2007, he was a Visiting Scholar with Waseda University. He is currently a Professor with Xiangtan University. He holds 13 invention patents. He has authored more than 30 articles. His main research interests include the Internet of Things, wireless sensor networks (WSNs), mobile ad-hoc networks, mobile communication networks, and social computing.



ZHIPING CHEN received the B.S. degree in computer science and technology from Xiangtan University, Xiangtan, China, in 1994, and the M.S. and Ph.D. degrees in computer science and technology from Hunan University, in 1997 and 2003, respectively. From 1997 to 2009, he was an Associate Professor with Hunan University. He is currently a Professor with Changsha University. His current research interest includes bioinformatics.

XUEYONG LI received the B.S. degree from Hunan Normal University, Changsha, China, in 1994, the M.S. degree in computer science and technology from Hunan University, in 2003, and the Ph.D. degree in computer science and technology from Northwestern Polytechnical University, in 2012. He is currently a Professor with Changsha University. His current research interest includes information theory.

WANG LEI received the Ph.D. degree in computer science from Hunan University, China, in 2005. From 2005 to 2007, he was a Postdoctoral Fellow of Tsinghua University, China. After that, he moved to Duke University, USA, and Lakehead University, Canada, as a Visiting Scholar. From 2009 to 2011, he was an Associate Professor with the College of Software, Hunan University. From 2011 to 2018, he was a Full Professor with the College of Information Engineering, Xiangtan University. He is currently a Full Professor and an Academic Leader of computer engineering with Changsha University, China. He has published more than 100 peer-reviewed articles. His main research interests include bioinformatics and the Internet of Things.