International Journal of Computer Applications (0975 - 8887)
Volume 186 - No.17, April 2024

Protein Function Prediction using Nearer Neighbor Proteins Interactions

Saima Khan
Faculty, Computer Science and Engineering
University of Development Alternative
Dhaka, Bangladesh

Saifuddin Md. Tareeq, PhD
Faculty, Computer Science and Engineering
University of Dhaka
Dhaka, Bangladesh

ABSTRACT

Protein is one of the most important elements of life. It is responsible for structuring organs, regulating various activities of human body, transporting materials throughout the body etc. Wet lab based experiments for determining protein functions are time consuming, difficult and expensive. Computational methods have got high demands in predicting the functions of proteins because these methods save a significant amount of time and are easier and less expensive than wet lab-based experiments. There are various approaches of predicting protein functions using computational methods. In this study, a novel idea has been proposed to predict the functions of proteins using protein-protein interaction network. This method is based on k-means clustering algorithm, nearer neighbor proteins functions and also the common neighbor proteins functions.

General Terms

protein, neighbor protein, prediction

Keywords

protein-protein interaction network (PPI), k-means clustering algorithm, support, common neighbor, second degree neighbor, protein function prediction

1. INTRODUCTION

A protein generally performs its functions associating with other proteins. Protein interactions with each other create a network of protein-protein interactions. A protein-protein interaction network can be presented by nodes and edges. Nodes represents proteins and edges represents the interactions among them. And this network is known as ”Protein-Protein Interaction Network”.
There are various methods for predicting protein functions. Among those methods, homology based methods were used highly. Homology-based methods annotate similar functions for proteins with similar structures. But these methods fail to predict protein functions when they find a protein of new structure or sequence. Protein function prediction using Protein Protein Interaction network has been popular for long time. Interacting proteins generally perform same types of functions. And based on this theory many methods have been proposed to predict protein functions. Many of them are complex method, time consuming and can’t provide higher accuracy. So, the aim of this study is to provide a simple computational method to predict protein functions.
Schwikowski et al. [17] Considers only direct interactions of proteins identified by biochemical experiments or two hybrid studies. For predicting function of a protein, they list all protein functions of its direct neighbors from highest frequent to lowest. Hishigaki et al. [3] developed an objective prediction method known as ”Neighborhood”, where the information of indirect interactions has been included. They have experimented for prediction of sub cellular localization, cellular role and the biochemical function of yeast proteins. Karaoz et al. proposed an approach (GenMulticut approach) [4] in which the functional linkage graph is mapped into a variant of the discrete-state Hopfield network. A district network is constructed for each function in GO (Gene Ontology). Each node is targeted to be assigned into one of three discrete states. Functional flow [7] considers each annotated node (protein) as a source of flow and considers each node without annotation as a recipient. Each edge has a specific capacity to carry the flow. For each protein, score of each function is obtained. This method incorporates the distance effect. The reservoir node that is nearer from the source, gets more flow than the nodes situated far away from the source node. For a particular protein, the function with the highest scores are assigned as its function. This method has considered all neighboring proteins and summed up the number of times each annotations occurs for each protein. For weighted graph, they have considered the weighted sum. Tania et al. [2] proposed a method that uses a minimum distance classifier to predict the function of unannotated protein. From the protein interaction network, hyper geometric distribution value and correlation coeficient of every protein have been calculated and used as features for this method. Though proteins are involved in so many functions for their work, only five functional groups (cell polarity, DNA repair, lipid metabolism, protein modification and protein synthesis) have been considered. Two different methods of experiments have been shown in their research paper. One is PFP_MINDSET1 and another method is PFP_MINDSET2. Peipei et al. [5] proposed a two node frequent pattern based method to predict function of unannotated protein on the basis of frequent pattern mining in graph data. This method is processed in 3 steps:
i. The first step is neighbor finding steps
ii. Second step is pattern finding
iii. The third step is function annotation

Chuan et al. [6] has proposed a novel common neighbor based model with Bayesian framework to predict protein function on the basis of small world property of protein-protein interaction network. According to this method, it is assumed that two proteins share common functions if they have more common neighbors between them.
To improve the accuracy of predicting protein functions, an innovative methodology has been developed in this study. Initially, the extensive protein network has been clustered into standardized components. Subsequently, a series of analyses have been conducted that leverage the functional attributes of neighboring proteins and the number of common neighbors. This approach enables to effectively predict protein functions in an easier and faster manner.

2. DATA

For this experiment (our experiment), two types of data have been used.
i. Protein-Protein Interaction Data
ii. Protein Function Data
These two types of data have been collected from string-db.org [8].

3. METHOD

There are two main parts in this method to predict protein functions. The first part is- "Clustering" and the second part is "Function Annotation".
i. Clustering: Clustering method is used to divide a dataset into the clusters in a way that the data points within each cluster are more alike to each other compared to those in other clusters.
ii. Function Annotation: "Function Annotation" part consists 3 steps within it. In the first step, it checks the functions of direct neighbors of the input protein and the counts the supports of the functions. Support indicates the frequencies of specific protein functions being present. Then, if necessary, this step weighs the functions and proceeds to the next steps. In the second step, it checks the functions of the second degree neighbors of the input protein, and in the third step, it counts the common neighbors between the input protein and its direct neighbor proteins, and then counts the supports of their functions.
Final function annotation is done based on the result of the integrated supports from these 3 steps.

3.1 Clustering

The protein-protein interaction network typically exhibits a complex structural organization, characterized by numerous edges denoting interactions among multiple proteins. However, not all edges within the network convey essential information regarding protein functionalities. Thus, the implementation of clustering algorithms emerges as a crucial step in the analytical process. Various clustering methodologies, such as K-means clustering, Markov clustering, Mean Shift clustering, and Hierarchical clustering, offer diverse strategies for delineating meaningful clusters within the network. For the present study, the K-means clustering algorithm has been selected due to its effectiveness in partitioning data. Initially, K-means clustering algorithm randomly selects centroids and assigns items to clusters based on proximity. This facilitates the aggregation of nodes with similar characteristics into same clusters. By focusing on clustered edges, which potentially harbor pertinent information about neighboring proteins, the analysis gains precision in delineating functional relationships. For instance, in a network comprising 51 proteins, an average of 571 edges may be present, yet not all contribute significantly to predicting protein functions. Consequently, post-clustering, attention is directed solely towards edges within clusters, thereby refining the predictive capabilities of the analysis. Figure 1 and figure 2 shows the condition of a graph before clustering and after clustering respectively.

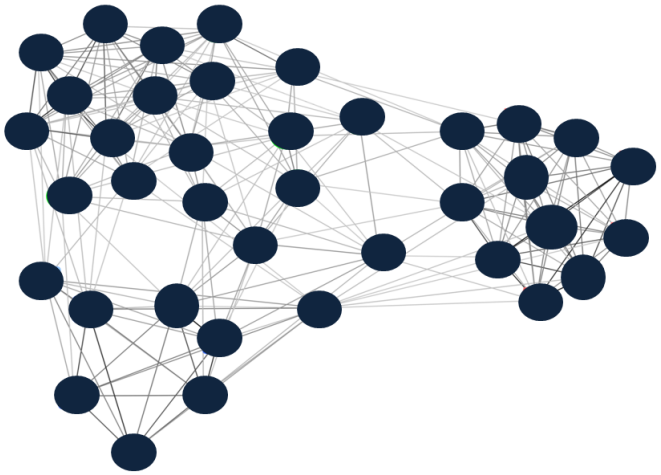

Fig. 1. A Sample Graph Without Clusters

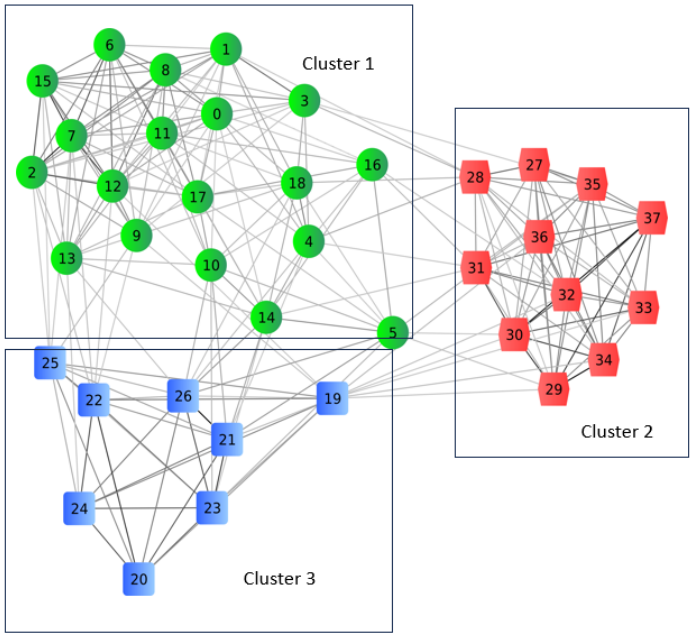

Fig. 2. The Graph After Clustering. Three rectangles indicate 3 different cluster groups

3.2 Annotation of Functions

For annotating function to the input protein, first the cluster of that input protein has to be identified. Then, within that cluster, 3 steps have to be followed but for some certain cases (explained in the description of Step 1), Only the first step is enough to effectively annotate the function of a protein. All the steps of the ”Function Annotation” have been described below.

3.2.1 Step 1. Figure 3 is considered as Case 1. In figure 3, the input protein is Y. The cluster of Y has to be considered and interactions within that cluster have to be considered for annotating functions. Y has 6 direct neighbor proteins and they are A, B, C, D, E and F. Among these 6 neighbor proteins, 3 proteins have the function f1, 2 proteins have the function f2 and 1 protein has function f3. So, here function f1 has the highest support and that is 3. So, the protein Y will be annotated with function f1 and no need to go to the next steps.

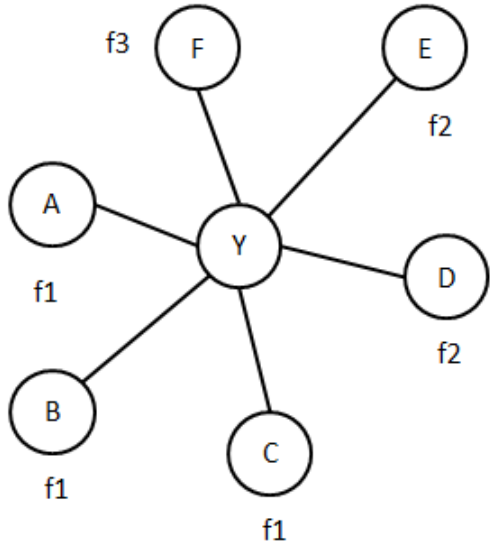

Fig. 3. Protein-Protien Interaction Network (Case 1)

Figure 4 is considered as Case 2. In figure 4, Z is the input protein. To annotate Z with an accurate function, the cluster of Z has to be viewed. In the cluster of Z, it has five direct neighbor proteins and they are a1, a2, a3, a4 and a5. Among those proteins, a1 and a2 have the function f1, a3 and a4 have the function f2 and a5 has the function f3. Function f1 and f2 both has the same highest support 2. As two distinct functions with identical supports have been identified at this stage, it is imperative to adhere to the instructions outlined in the subsequent steps to ascertain the particular higher dominating function. Before going to the next steps, the supports of the functions have to be provided an equal higher weight. In this context, the supports of the functions associated with the direct neighbor proteins of protein Z have been assigned a weight that is twice their original value. Table 1 presents the direct neighbors of the input protein Z and their respective functions. Table 2 and table 3 shows the weightless and weighted supports of the direct neighbor proteins' functions of protein Z for Case 2.

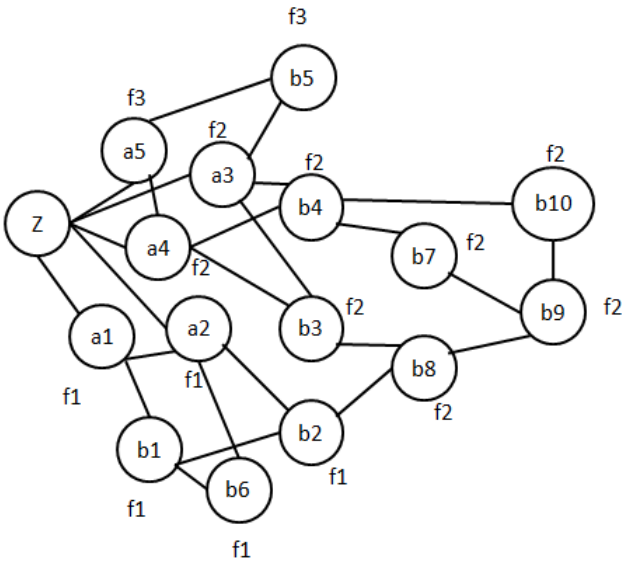

Fig. 4. Protein-Protien Interaction Network (Case 2)

Table 1. The Direct Neighbor Proteins and Their Functions (Case 2)

Proteins	Functions
a1	f1
a2	f1
a3	f2
a4	f2
a5	f3

Table 2. Functions with actual support from step 1 (Case 2)

Functions	Actual Supports
f1	2
f2	2
f3	1

3.2.2 Step 2. This step involves examining the supports of the functions of the neighbor proteins located at a distance of two units (two radius) away from the input protein Z (For Case 2). For performing this step, consider the direct neighbors of a1, a2, a3, a4 and a5 from figure 4. These (a1, a2, a3, a4 and a5) proteins are the direct neighbors of the input protein Z. The direct neighbors (excluding Z) of proteins a1, a2, a3, a4, and a5 are situated at a distance of two units from the input protein Z. The direct neighbors of these five proteins can be referred to as ”Second-degree Neighbor Proteins of Input Protein Z”. Table 4 presents the proteins located at a distance of 2 units from the input protein Z, serving as its second-degree neighbors.

3.2.3 Step 3. The task of this step is to find the number of common neighbors of input protein Z and its' direct neighbors (a1, a2, a3, a4 and a5). In this step, supports are given according to the number of common neighbors the input protein shares with its di-

Table 3. Functions with weighted supports from Step 1 (Case 2)

Functions	Weighted Supports
f1	4
f2	4
f3	2

Table 4. Functions of the proteins located at a distance of 2 units from the input protein Z, serving as the second-degree neighbors of protein Z (From Step 2 in Case 2)

Protein	Their Direct Neighbors (and 2nd Degree Neighbor of Z)	Function
a1	a2	f1
a1	b1	f1
a2	a1	f1
a2	b2	f1
a2	b6	f1
a3	b3	f2
a3	b4	f2
a3	b5	f3
a4	b3	f2
a4	b4	f2
a4	a5	f3
a5	a4	f2
a5	a5	f3

Table 5. Supports of the Functions from Step 2

Function	Supports
f1	5
f2	5
f3	3

rect neighbors. Table 6 shows the number of common neighbors and their functions.
From table 6, we see that function f1 is present twice and f2 and f3 is present once. so, the respective support of f1 is 2, f2 is 1 and f3 is also one.
After completion of three steps, supports from these 3 steps have to be integrated. The supports have been integrated from table 3, 5, and 7. Table 8 shows the integrated result of the supports of the functions. From table 8, it is shown that function f1 has the highest integrated support. So, Protein Z will be annotated with function f1 according to this study. If there are multiple functions with the same highest supports after this final calculation, multiple functions can be assigned for the input protein.

4. RESEARCH WORK AND OUTCOME

This work has been performed on a protein interaction data (of Saccharomyces cerevisiae) which has been collected from STRING [8]. This network contains 1224 proteins and 13704 edges. K-means clustering algorithm has been used to cluster this large PPI(Protein Protein Interaction) network. After clustering, edges outside the clusters have not been considered for the PPI interaction and function prediction. Only edges within the clusters have been considered for the experiment. Leave-one-out cross validation has been performed for this experiment. Leave-one-out cross-validation is like taking K-fold cross-validation to the maximum-here, K equals the total number of data points (N) in the set. So,

Table 6. Common neighbors of input protein Z and its direct neighbor proteins (Step 3)

Pairs	Common Neighbors	Total Common Neighbors	Functions
$(Z, a1)$	a2	1	f1
$(Z, a2)$	a1	1	f1
$(Z, a3)$	-	0	-
$(Z, a4)$	a5	1	f3
$(Z, a5)$	a4	1	f2

Table 7. Supports of Functions from Step 3

Function	Support
f1	2
f2	1
f3	1

Table 8. Integrated Supports of the Functions

Function	Integrated Support
f1	(4+5+2)=11
f2	(4+5+1)=10
f3	(2+3+1)=6

the function approximator gets trained N times, each time on all the data except for one point, and then predicts for that left-out point.The experiment has been conducted in 12 parts.
The experiment has been conducted by implementing the method using C programming language within the MATLAB environment. Function prediction for molecular function and biological process have been shown in table 9 and table 10 respectively.
The metrics- sensitivity, specificity, and accuracy have been used to analyze the result.
TP: Number of Positive Samples that predicted as positive
TN: Number of negative Samples that predicted as negative
FP: Number of negative Samples that predicted as positive
FN: Number of positive Samples that predicted as negative

Sensitivity: It is the probability of correct prediction of a class.

$$sensitivity = \frac{TP}{P}$$

Specifity: It is the probability that a positive prediction for the class is correct.

$$Specificity = \frac{TN}{N}$$

Accuracy: It is the overall probability that the prediction is correct.

$$Accuracy = \frac{TP + TN}{P + N}$$

The result analysis have been performed in two ways. The first way provides the overall performance of the result (in table 9, table 10, and in figure 5). In the second way, only the most frequent five functional groups have been considered and the performance have been shown with the scores of sensitivity, specificity and accuracy (table 11).
This work has been compared with two existing research works. One is Two-Node frequent patterns [22] and another is

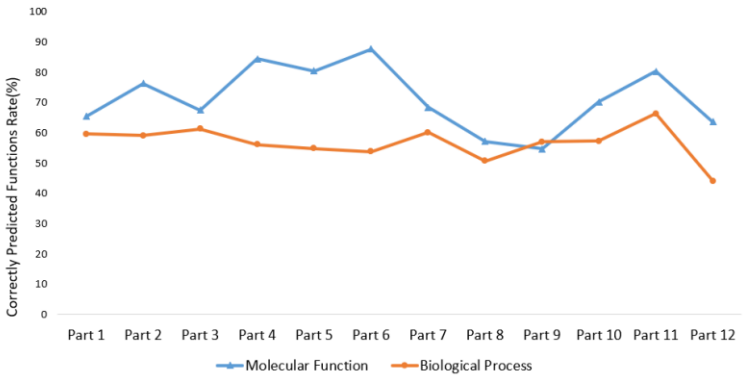

Fig. 5. Accuracy of the Prediction Rate (%) of Molecular and Biological Functions

Table 9. Accuracy of Molecular Function Prediction By the Proposed Method

Part	Correctly Predicted Functions
Part 1	65.5%
Part 2	76.3%
Part 3	67.5%
Part 4	84.5%
Part 5	80.5%
Part 6	87.7%
Part 7	68.5%
Part 8	57.2%
Part 9	54.7%
Part 10	70.3%
Part 11	80.3%
Part 12	63.7%
Average	71.39%

Table 10. Accuracy of Biological Process Prediction By the Proposed Method

Part	Correctly Predicted Functions
Part 1	59.6%
Part 2	59.1%
Part 3	61.3%
Part 4	56.2%
Part 5	54.9%
Part 6	53.9%
Part 7	60.2%
Part 8	50.8%
Part 9	57.1%
Part 10	57.3%
Part 11	66.4%
Part 12	44.1%
Average	57.57%

PFP_MINDSET1 [23]. For comparing with two node frequent patterns, the overall accuracy of molecular function prediction and biological process prediction has been considered. Table 12 and figure 6 presents the comparison of predicting accurate molecular functions between the proposed method and Two Node Frequent Pattern. Table 13 and figure 7 presents the comparison

Table 11. Result Analysis 2: In this result analysis, the whole functional groups have been divided into 5 most frequent and dominant molecular functional categories. For each of these categories, sensitivity, specificity and accuracy have been measured. Here, 73.4% accuracy (overall accuracy) has been obtained.

Functional Group	Sensitivity	Specificity	Accuracy (%)
Hydrolase	0.85	0.877	86.7%
Oxidoreductase	0.714	1	80%
Transferase	0.69	0.8	72%
Lyase	0.67	0.60	69%
Ligase	0.8	0.57	60%

of predicting accurate biological processes between the proposed method and Two Node Frequent Pattern. And for comparing with the PFP_MINDSET1, only the 5 most frequent functional categories have been chosen to measure and compare the sensitivity, specificity and accuracy. Comparison 1 shows that the proposed method performs better than "Two-Node Frequent patterns" [22] and comparison 2 (table 14) shows that the performance of PFP_MINDSET1 [23] and our proposed method is nearly same.

5. CONCLUSION

Proteins are essential biomolecules with diverse functions crucial for life processes. They provide structural support to cells, tissues, and organs, and also play key regulatory roles in modulating physiological processes. Knowing protein functions is fundamental for unraveling the complexities of biological processes and optimizing therapeutic strategies. This understanding helps us see the complicated ways that metabolism and disease pathways work, making it easier to create and improve personalized drugs.

Many protein functions remain unknown, and traditional wet-lab experiments are characterized by their time-consuming, labor-intensive nature, and high cost, necessitating significant human resources. In response, scientists and researchers are actively endeavoring to innovate more efficient protein function prediction methods utilizing computational approaches. Towards this end, a straightforward computational method has been developed here which aims at accurately predicting protein functions. This method employs the k-means clustering algorithm to organize protein-protein interaction networks, subsequently annotating functions to proteins of unknown function based on higher-order functions

Table 12. Comparison of Molecular Function Prediction With Two-Node Frequent Pattern (Comparison 1)

Part	Partly Accuracy of the Proposed Method	Partly Accuracy of Two-Node Frequent Pattern
Part 1	65.5%	62.3%
Part 2	76.3%	71.5%
Part 3	67.5%	63.7%
Part 4	84.5%	82.5%
Part 5	80.5%	75.3%
Part 6	87.7%	84.5%
Part 7	68.5%	63.7%
Part 8	57.2%	54.3%
Part 9	54.7%	50.4%
Part 10	70.3%	69.5%
Part 11	80.3%	73.3%
Part 12	63.7%	56.5%
Average	71.39%	67.3%

Table 13. Comparison of Biological Process Prediction With Two-Node Frequent Pattern (Comparison 1)

Part	Partly Accuracy of the Proposed Method	Partly Accuracy of Two-Node Frequent Pattern
Part 1	59.6%	50.3%
Part 2	59.1%	53.6%
Part 3	61.3%	49.8%
Part 4	56.2%	52.3%
Part 5	54.9%	48.7%
Part 6	58.9%	56.5%
Part 7	60.2%	47.3%
Part 8	50.8%	43.3%
Part 9	62.1%	54.3%
Part 10	58.3%	50.6%
Part 11	66.4%	64.7%
Part 12	44.1%	45.5%
Average	57.57%	51.41%

and shared neighbor proteins. Demonstrating efficacy, this method offers a streamlined approach that minimizes time investment while enhancing accuracy compared to several established protein function prediction methods. When there aren't enough well-understood proteins within 2 units of the nearest neighbors from the protein being studied, the effectiveness of this research approach might suffer. This happens because this study relies on nearby proteins to understand the functions, and they need to be close in the protein-protein interaction network for accurate predictions. Trying to expand this method beyond two levels of nearest neighbors could make it less accurate. This is because including proteins that are far away could add irrelevant information, making the predictions less accurate. So, it's important to be careful about how far we look to maintain the reliability of the predictions. However, even with this limitation, this method is still useful for predicting protein functions within a limited network. Researchers should just be cautious when using it and explore other ways to predict functions for proteins further away from the one being studied.

Creating an effective global protein function prediction method requires a systematic approach that integrates various types of data, such as protein sequence and structure data, gene expression data, pathway analysis, and protein-protein interaction data. Different methodologies like clustering, using neighbor proteins information, and association analysis-based approaches can be employed in this process. By using ensemble classifiers that combine information from gene expression data, protein-protein interaction data, and protein sequence data, we can enhance the accuracy of protein function prediction methods. Additionally, improving the clustering algorithm is crucial for enhancing the precision of protein function prediction. Advancements in clustering techniques offer potential for refining the accuracy and reliability of protein function prediction methods, thus contributing significantly to progress in biological research and drug discovery efforts.

6. REFERENCES

[1] Emrah Akkoyun and Tolga Can. "Parallelization of the functional ow algorithm for prediction of protein function using protein-protein interaction networks". In: (2011), pp. 56-62.

[2] Tania Chatterjee, piyali Chatterji, Shubhadip Basu, Mahantapas Kundu, Mita Nasipuri, "Protein function by minimum distance classifier from protein interaction network", 2012, pp. 588-591.

[3] Haretsugu Hishigaki, Kenta Nakai, Toshihide Ono, Akira Tanigami, Toshihisa Takagi, "Assessment of prediction accuracy of protein function fromproteinprotein interaction data". In: Yeast 18.6 (2001), pp. 523-531.

[4] Ulas Karaoz, T. M. Murali, Stan Letovsky, Yu Zheng, Chunming Ding, Charles R. Cantor, and Simon Kasif, "Whole-genome annotation by using evidence integration in functionallinkage networks". In: Proceedings of the National Academy

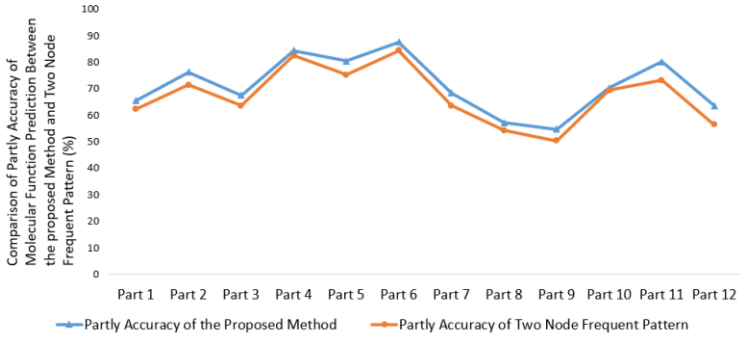

Fig. 6. Comparison of Accurately Predicting Molecular Functions by the Proposed Method and Two Node Frequent Pattern(%)

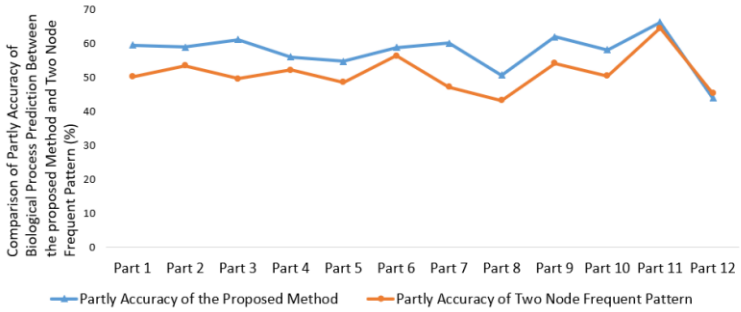

Fig. 7. Comparison of Accurately Predicting Biological Processes by the Proposed Method and Two Node Frequent Pattern(%)

of Sciences of the United States of America 101.9 (2004), pp. 2888-2893.

[5] Peipei Li, Lyong Heo, Meijing Li, Keun Ho Ryu, Gouchol Pok, ,"Protein function prediction using frequent patterns in protein-protein interaction networks", 2011, pp. 1616-1620.

[6] Chuan Lin, Daxin Jiang, and Aidong Zhang. "Prediction of protein function using common neighbors in protein-protein interaction networks". In: (2006), pp. 251-260.

[7] Elena Nabieva, Kam Jim, Amit Agarwal, Bernard Chazelle and Mona Singh, "Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps". In: Bioinformatics 21.suppl 1 (2005), pp. i302-i310.

[8] String. url: string-db.org.

[9] Uniprot. url: uniprot.org

[10] Arvind Kumar Tiwari and Rajeev Srivastava, "A survey of computational intelligence techniques in protein function prediction". In: International journal of proteomics 2014 (2014).

[11] Alexei Vazquez, Alessandro Flammini, Amos Maritan, Alessandro Vespignani. "Global protein function prediction from protein-protein interaction networks". In: Nature biotechnology 21.6 (2003), pp. 697-700.

[12] Wei Xiong, Hui Liu, Jihong Guan, Shuigeng Zhou, "Protein function prediction by collective classification with explicit and implicit edges in protein-protein interaction networks". In: BMC bioinformatics 14.12 (2013), S4.

[13] Dengdi Sun, Maolin Hu, "Determining Protein Function by Protein-protein Interaction Network", 2007, IEEE.

[14] Young-Rae Cho, Aidong Zhang, Discovering frequent patterns of functional associations in protein interaction networks for function prediction, IEEE International Conference on Bioinformatics and Biomedicine, 2008.

[15] Qingshan Ni1, Zhengzhi Wang1, Qingjuan Han2, Gangguo Li1, Xiaomin Wang1, Guangyun Wang1, "Using logistic regression method to predict protein function from protein-protein interaction data", IEEE, 2009.

[16] Young-Rae Cho, Member, IEEE, and Aidong Zhang, "Predicting Protein Function by Frequent Functional Association Pattern Mining in Protein Interaction Networks", IEEE, 2008.

[17] Lei Shi, Young-Rae Cho, Aidong Zhang, "ANN Based Protein Function Prediction Using Integrated Protein-Protein Interaction Data", International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009.

[18] Benno Schwikowski, Peter Uetz, and Stanley Fields. "A network of proteinprotein interactions in yeast". In: Nature biotechnology 18.12 (2000), pp. 1257-1261.

[19] Guoxian Yu, Huzefa Rangwala, Carlotta Domeniconi, Guoji Zhang, and Zhiwen Yu, "Protein Function Prediction with Incomplete Annotations", IEEE/ACM Transactions On Computational Biology and Bioinformatics, Vol. 11, NO. 3, May/June 2014.

[20] Wei Peng, Min Li , Lu Chen, and Lusheng Wang, "Predicting Protein Functions by Using Unbalanced Random Walk Algorithm on Three Biological Networks", IEEE/ACM Transactions On Computational Biology and Bioinformatics, Vol. 14, NO. 2, March/April 2017.

Table 14. Comparison with PFP_MINDSET1 (Comparison 2)

Functional Group	Proposed Method			PFP_MINDSET1		
	Sensitivity	Specificity	Accuracy (%)	Sensitivity	Specificity	Accuracy (%)
Hydrolase	0.85	0.877	86.7%	0.968	0.8	86.2%
Oxidoreductase	0.714	1	80%	0.975	0.826	84.2%
Transferase	0.69	0.8	72%	0.869	0.727	79.8%
Lyase	0.67	0.60	69%	0.843	0.719	74%
Ligase	0.8	0.57	60%	0.716	0.798	58%

[21] Sovan Saha, Piyali Chatterjee, Subhadip Basu, Mahantapas Kundu, Mita Nasipuri, "Improving prediction of protein function from protein interaction network using intelligent neighborhood approach.", Communications, Devices and Intelligent Systems (CODIS), 2012 International Conference on, pages 584–587, IEEE, 2012.

[22] Peipei Li et al. "Protien Function Prediction Using Frequent Patterns in Protein-Protein Interaction Network". In: (2011). doi: 978-1-61284-181-6/11/$26.00

[23] Tania Chatterjee Li and Piyali Chatterjee. "Protien Function Prediction by Minimum Dis- tance Classifier from Protien Interaction Network". In: (2012). doi: 978 - 1 - 4673 - 4700 - 6/12/$31.00.

[24] Hon Nian Cha, Gumei Liu, Limsoon Wong, "Protein function prediction using protein–protein interaction networks". Protein Function Prediction for Omics Era (2011), pp.243-270.

[25] Sobhan Moosavi, Masoud Rahgozar, Amir Rahimi, "Protein function prediction using neighbor relativity in protein–protein interaction network". Computational Biology and Chemistry Volume 43, Pages 11-16 (April 2013).