Protein Complexes Detection Based on Node Local Properties and Gene Expression On PPI Weighted Networks

Yang Yu (yuyangsd1204@126.com)
Software College, Shenyang Normal University

Dezhou Kong
Software College, Shenyang Normal University

Research Article

Keywords: protein complex, protein-protein interaction (PPI), node local properties ,weighted graph construction.

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-287016/v1

Version of Record: A version of this preprint was published at BMC Bioinformatics on January 6th, 2022.
See the published version at https://doi.org/10.1186/s12859-021-04543-4.

Protein Complexes Detection Based on Node Local Properties and Gene Expression On PPI Weighted Networks

Yang Yu Dezhou Kong

Software College, Shenyang Normal University, Shenyang, P.R. China

Abstract

Background

Identifying protein complexes from protein–protein interaction (PPI) networks is a crucial task, and many related algorithms have been developed to solve this issue. These algorithms usually consider a node's direct neighbors and ignore resource allocation and second-order neighbors. The effective use of such information is crucial to protein complex detection.

Results

To overcome this deficiency, this paper proposes a new protein complex identification method based on node-local topological properties and gene expression information on a new weighted PPI network, named NLPGE-WPN (joint node-local topological properties and gene expression information on weighted PPI network). First, based on the resource allocation of the PPI network and gene expression, a new weight metric is designed to describe the interaction between proteins. Second, our method constructs a series of dense complex cores based on density and network diameter constraints; the final complexes are recognized by expanding the second-order neighbor nodes of core complexes. Experimental results demonstrate that this algorithm has improved the performances of precision and f-measure, which is more valid in identifying protein complexes.

Conclusions

This identification method is simple and can accurately identify more complexes by integrating node-local properties and gene expression on PPI weighted networks.

Keywords: protein complex, protein-protein interaction (PPI), node local properties ,weighted graph construction.

E-mail:yuyangsd1204@126.com

Background

Proteins are the basis of biological activities, and their functions are generally expressed by the interactions between proteins [1]. In organisms, protein-protein interaction (PPI) networks consist of proteins and protein interactions. PPI networks provide an elegant means for expressing gene regulation and metabolic pathways in complex biological systems [2]. Protein complexes are the locally dense regions of PPI networks and possess graph-like structures in which a node represents a protein and an edge represents interaction between two proteins[3]. Complexes take part in many diverse biochemical activities that are fundamental to all kinds of functions, such as cell homeostasis, cell cycle control, growth, and proliferation. Moreover, specific functional modules usually are related to certain diseases.

Although great progress has been made in identifying protein complexes, laboratory-based methods are expensive, ineffective and sometimes even infeasible, and only parts of protein complexes are located. In addition, experiments in the laboratory are often incomplete because of the constraints of experimental conditions. As it is necessary to overcome the lacking of laboratory-based methods, a large number of computational algorithms have been designed as alternative methods to identify protein clusters, such as density-based clustering [4-6], hierarchical clustering [7,8], partition-based clustering [9,10], flow simulation-based clustering [11-13] and other methods [14-21]. In order to combine multiple information and mine more biological complexes, the source of additional biological information has recently been used to locate protein complexes in PPI networks [22-24].In fact, edges or vertices often contain some important potential information in PPI networks. Most scholars consider the definition of the weight of the edges in the network based solely on the topology, or only consider the realistic edge weights in the network for community discovery. How to employ and combine the

biological edge weight with the network topology as the base of the community discovery algorithm, thereby finding more overlapping complexes in the PPI network, is a topic worthy of further study.

Although the mining of protein complexes based on PPI networks has achieved some results, how to reasonably integrate PPI node data and gene expression biological information to construct weighted graphs, and how to define effective detection methods to identify complexes from the weighted network still need further study. In addition, only direct neighbors are applied to PPI network clustering problems, which is not sufficient. Aiming at a solution for the above-mentioned problems, a new graph clustering algorithm (called NLPGE-WPN) is proposed for protein complex detection on a new weighted graph and core-attachment structure. First, based on the resource allocation and gene expression of the PPI network, a new weight metric is designed to accurately describe the interaction between proteins. Then our method constructs a series of dense complex cores based on density and network diameter constraints; the final complexes are recognized by expanding the second-order neighbors of nodes in core complexes. This identification method is simple and can accurately identify more complexes. Experimental results demonstrate that this algorithm has improved performance in terms of precision and the f-measure, more effectively identifying protein complexes.

RESULTS

Datasets

CYC2008 [29] and MIPS [30] are used as benchmark complex datasets to assess the performance of different methods, and protein complexes with sizes longer than 3 are used. The

PPI data of S.cerevisiae is from DIP [31]. GSE3431 dataset [32] is employed in our paper, which records the data of 36 time points during three successive metabolic cycles.

Evaluation Criteria

In this section, the measure of matching rule is presented to assess how well a located cluster matches a real complex. Given a set of real protein complexes $R=\{R_1,R_2,...,R_n\}$ and a set of identified clusters $P=\{P_1,P_2,...,P_n\}$, a identified protein cluster P_i is supposed to match a real protein complex R_j if ω value is not less than 0.2. In order to evaluate the performance among the different methods, recall, precision and f-measure [33] are utilized as follows.

$$\omega\frac{|C|^2}{|P_i|*|R_j|} \tag{1}$$

$$recall=\frac{|\{R_j \mid R_j\in R\wedge\exists P_i\in P, P_i matches R_j\}|}{|R|} \tag{2}$$

$$precision=\frac{|\{P_i \mid P_i\in P\wedge\exists R_j\in R, R_j matches P_i\}|}{|P|} \tag{3}$$

$$f-measure=\frac{2*recall*precision}{recall+precision} \tag{4}$$

Here, $|C|$ is the number of common proteins in the interaction set between a located cluster P_i and a real complex R_j. $|P_i|$ and $|R_j|$ are the numbers of proteins in P_i and R_j, respectively.

Comparison With Other Methods

To assess the performance of our proposed algorithm,we compare our algorithm with MCODE [6], CFinder [34], ClusterONE [17], MCL [35], L [7] and PEWCC [36] in terms of precision, recall and f-measure in the DIP PPI network.

From Fig.1, it is observed that both the precision value and the f-measure value are higher than other methods when the α parameter value is set to 0.3 on data set CYC2008. The recall values of L and PEWCC algorithms are superior to our method, which are 0.615 and 0.696, respectively. From Fig 2, our method can achieve the best f-measure value among seven methods and the recall value is higher than CFinder,MCODE,MCL,L,ClusterONE except L and PEWCC. Through comparison and analysis, we found that MCODE, CFinder, ClusterONE and L only employ the topological features of a single network for protein complex identification, so relatively few indicators can be selected and employed in complex detection and influence the performance of protein complexes identification. The NLPGE-WPN algorithm proposed in this paper combines PPI network local nodes characteristics and gene expression data to solve overlapping protein complex discovery problems. Furthermore, experimental results show the f-measure value of our method is higher than other typical algorithms' f-measure value. It indicates that the performance of NLPGE-WPN algorithm is optimal.

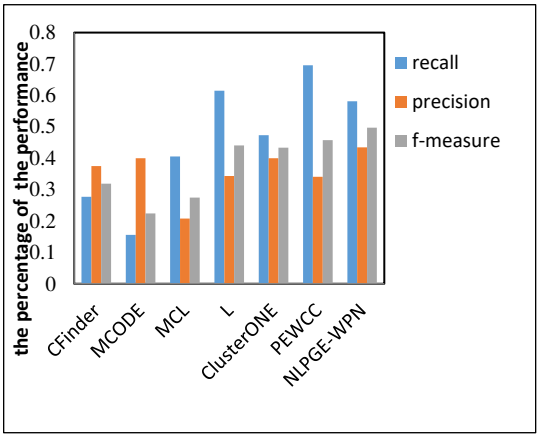

Fig.1. Comparison On Dataset CYC2008

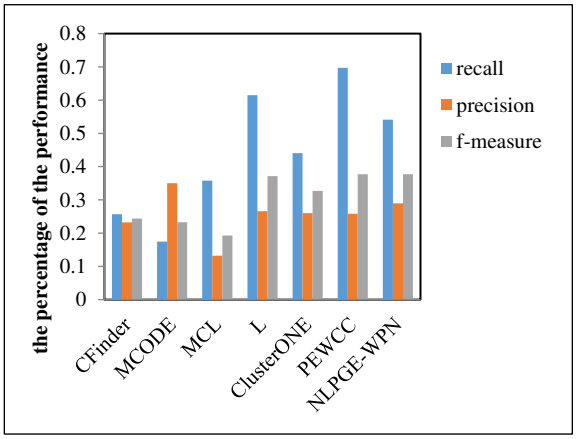

Fig.2. Comparison On Dataset MIPS

Assessment of the Importance for Key Parameter α on Clustering Performance

By evaluating the importance of key parameters, we can more intuitively observe the influence of a certain parameter on the experimental results, and it is helpful to understand the advantages and disadvantages of the algorithm and enhance it.

The critical parameter α in our method is mainly employed to show the effectiveness of information fusion from local neighbors and gene expression information and to affect the detection results of protein complexes. The value range of the parameter α is [0,0.9]. This experiment investigates the effects of different values from 0.1 to 0.9 at intervals of 0.1 on complex detection performance. Using f-measure as our experimental evaluation criterion, the performances of the key parameter are evaluated as shown in the Fig.3. It can be seen that when the critical parameter α is greater than or equal to 0.3, the f-measure tends to stable. In this article, we take $\alpha = 0.3$.

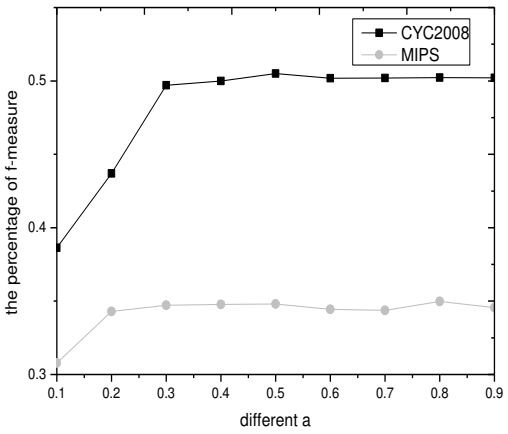

Fig .3 The influence of key parameters α on experimental results of f-measure

Robustness to the Different Thresholds (t)

In order to illustrate the comprehensive performance of NLPGE-WPN, we demonstrate f-measure performances with nine thresholds t=$\{0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9\}$ among

different methods in Fig.4. Fig.4(a) shows the f-measure performances on the CYC2008 benchmark dataset with other six methods.It can be inferred that NLPGE-WPN performs better than CFinder,MCODE,MCL,L, ClusterONE and PEWCC. Similar results can also be inferred on the MIPS benchmark in Fig.4(b). By reaching an average level of recall and precision, our method can get better performance from the PPI network to detect protein complexes on both complex data sets. This further demonstrates the effectiveness of the fusion information from local node and gene expression data.

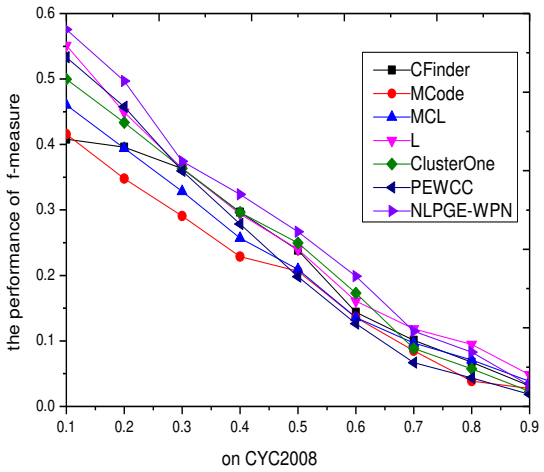

Fig .4(a) F-measure comparisons on CYC2008

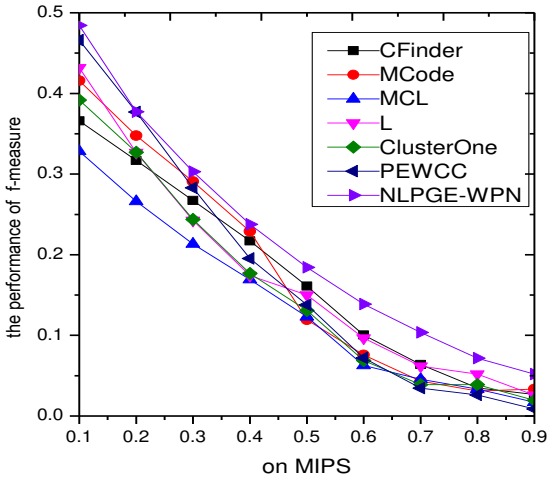

Fig .4(b) F-measure comparisons on MIPS

Functional Analysis

For the protein complexes identified by the NLPGE-WPN algorithm, we must measure not only the effectiveness of the algorithm in terms of quantity, but also qualitatively. We analyze the biological significance of the identified protein complexes. Real protein complexes often present high functional homogeneity, so the function enrichment test is employed to demonstrate in our paper the biological significances of detected protein complexes. P-value is the statistical significance of the occurrence of a complex if the p-value is less than 0.01. The functional homogeneity of a located cluster is usually the smallest p-value over all the possible functional

clusters. We present some examples of predicted complexes from CYC2008 dataset in Table 1 with their p-values. Moreover, the function of unknown proteins from the located clusters can be predicted because of proteins in the same complex with the same function. Based on this observation, it is useful to infer the type of protein function. As shown in Fig. 5, the complex with size 7 is predicted in our experiment and 6 proteins are treated as tRNA wobble uridine modification. It can be inferred that an uncharacterized protein YOR151C can have the same function as tRNA wobble uridine modification. It also proves that node YOR151C and node YLR384C have no common neighbors, but YOR151C still can transmit information to YGR200C, YPL101W, YPL086C, YHR187W and YMR312W via node YLR384C. It further shows the importance of the use of RA.

Table 1 Located protein complexes and their p-values

Gene name	p-values	Main functional group
YGL112C YBR198C YGR274C YER148W YDR448W YDR167W YMR236W YDR176W	2.05e-12	histone acetylation
YOL021C YOR076C YDL111C YGR095C YNL189W YNL232W YGR195W YGR158C YHR069C YDR280W YCR035C	4.96e-25	nuclear-transcribed mRNAcatabolic process, exonucleolytic, 3 '-5'
YDR473C YMR213W YLR147C YLR117C YHR165C YKL173W YER172C YJR022W YGR091W YBR055C YLR438C-A YPR178W	1.30e-20	mRNA splicing, via spliceosome
YFL009W YJR033C YDR328C YDL132W YOL133W YDR202C	6.13e-08	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process
YHR052W YPL043W YNL061W YOR061W YGR090W YIL035C YGR103W YOR039W YPR016C YGL019W YHR066W YNL132W YMR049C YJL069C	1.36e-08	rRNA processing
YGL127C YOL135C YER022W YBL093C YGR104C YHR058C YDL005C YBR193C YPL042C YHR041C YBR253W YOL051W YPL248C	6.09e-14	positive regulation of transcription by RNA polymerase II

YPR072W YCR093W YDL165W YNR052C YER068W YAL021C YNL288W YIL038C YDR443C	3.69e-15	regulation of transcription elongation from RNA polymerase II promoter

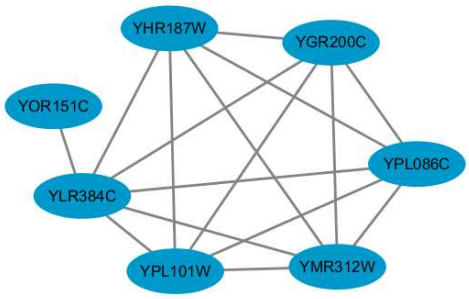

Fig.5 One example of predicted complex

Conclusions

The identification of protein complexes is important for discovering and understanding the cellular organizations and biological processes in PPI networks. In this paper a new approach named NLPGE-WPN is proposed for identifying protein complexes in protein-protein interaction networks. Based on the resource allocation and gene expression of the PPI network, we first design a new weight metric to accurately describe the interaction between proteins. Our method then constructs a series of dense complex cores based on density and network diameter constraints, and the final complexes are recognized by expanding the second-order neighbors of nodes in core complexes. This identification method is simple and can accurately identify more complexes. Experimental results demonstrate that this algorithm has improved performance as to precision and f-measure, making it valuable in identifying protein complexes. We hope our work may help the bioinformatics researcher to find more undiscovered protein complexes.

Methods

Protein complex detection with a computational approach from PPI data is useful as a supplement to the limited experimental methods. Besides the enhancement in graph clustering techniques, successful and accurate protein complex prediction also depends on the construction of weighted graphs. The noise of the protein interaction data is still an important factor to influence the performance of computational methods. Therefore, evaluating the reliability of protein interactions is essential. In this section, we introduce a novel method based on resource allocation and gene expression in weighted PPI networks (called NLPGE-WPN) which has two main steps. First, a new method is proposed to evaluate the reliability of the protein interaction data considering the common neighbor information and gene expression profiles and a new weighted graph is constructed. Second, protein complexes are detected based on core-attachment method in this new weighted graph. The workflow of our method is shown in Fig.6.

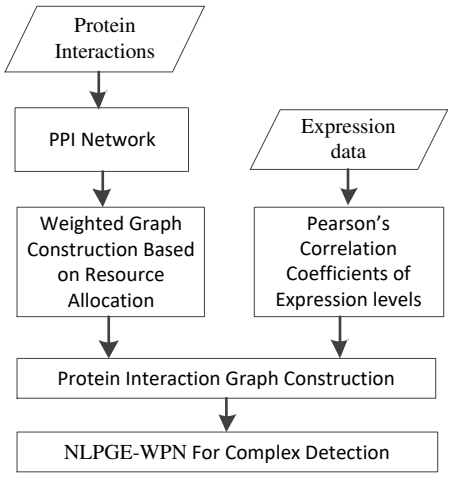

Fig.6. The workflow of our method

Assessing the Reliability of Protein Interaction

In this section, we design a new method to measure the reliability of protein pairs. To represent a PPI network, a 3-element tuple $G = (V, E, W)$ is employed, where $V = (V_i)(1 \le i \le N)$ is a set of N proteins, and $E = \{e_{ij}\}$ is the set of PPI edges whose value is M. For each pair of nodes, $i, j \in V$ and the edge e_{ij} is assigned a score as w_{ij}. Prompted by the reference [25], we introduce a resource allocation index, RA, to measure the similarity of interaction proteins. The concept of community structure shows that if two nodes have more than one common neighbor, then the relationship between them is close, and naturally they are more likely to belong to the same community. Similarity quantity can be used to evaluate the similarity between two nodes in a network. Therefore, a weighted graph based on resource allocation (WRA) is used in this paper to evaluate the weighted edge between them.

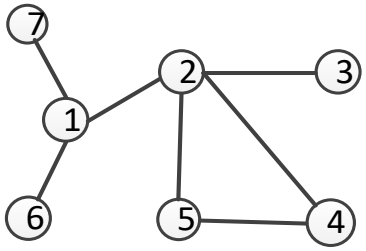

Fig.7. Sample network

Taking Fig.7 as an example, there is an edge between node 1 and node 2 and no common neighbors between them, but e_{12} is an important bridge for information transmission between node group{6, 7} and node group{3, 4, 5}. Simply, it is assumed that each transmitter carries a unit of resource, and will equally deliver it among all its neighbors. Based on this, we introduce the similarity [26] of two nodes in Equation (5). We can consider i and j, which are not directly

connected, but the node i still can transmit the information to node j with their common neighbors. WRA measures require only the information of the nearest neighbors which therefore have very low computational complexity. $N(i)$ is the neighbors of node i and N(j) is the neighbors of node j.

$$WRA_{ij} = \sum_{u \in N(i) \cap N(j)} \frac{1}{N(u)} \quad (5)$$

$$W_N = \{WRA_{ij}\} \quad (6)$$

Pearson's Correlation of Expression Levels

It is found that co-expression genes tend to encode interacting proteins [27]. For interacting protein pair p and q, Pearson's correlation coefficient of expression levels(PCC) is calculated. A higher correlation suggests a higher confidence in their interaction. PCC is generally used to measure the strength of the linear relationship between two variables and is also commonly used to measure the linear relationship between two sets of gene expression values. Suppose there are two columns of gene expression profiles $X=(x_1,...,x_n)$ and $Y=(y_1,...,y_n)$. The PCC calculation formula is defined in Equation (7)

$$PCC_{ij} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (7)$$

$$W_p = \{PCC_{ij}\} \quad (8)$$

where $\bar{x}$ denotes the average value of the expression value of gene X at 36 different times and $\bar{y}$ denotes the average value of the expression value of gene Y at 36 different times. In our study, we will calculate the value between each pair of interacting proteins as part of the weighted value.

Weighted Graph Construction

In this part, we will first describe how to compute the weighted value by combining gene expression information (GEI) based on PCC and RA information between two interaction proteins. A new weighted construction formula is proposed in Equation (9).

$$W = \alpha W_N + (1-\alpha) W_P \tag{9}$$

PCC represents pearson correlation coefficient and WRA for weighted graph construction, respectively. The value range of constant α is [0,1]. When $\alpha = 1$, the weighted method only considers RA information. When $\alpha = 0$, the weighted method only considers gene expression information. According to GBA principle (i.e. genes with similar expression spectrums have similar biological functions) [28] and the definition of RA, the weighting method of Equation (9) can better measure the differential importance of interaction in protein networks so as to improve the reliability of protein network as a whole. Next, we use Equation (9) to construct a new protein weighted network.

Detecting Protein Complexes in Weighted Graphs

The proposed algorithm, NLPGE-WPN, consists of two phases: weighted graph construction and core-attachment protein complex detection based on second-order neighbors. In the weighted graph construction phase, gene expression information and common neighbor information are integrated. A detailed description of the algorithm is outlined in Algorithm 1. Line 1 is for constructing the weighted network W_N with the given PPI dataset. Line 2 is for constructing the weighted network W_{p} with the gene expression data. Line 3 is for constructing the new weighted graph W based on W_N and W_{p}, and the protein interaction confidence is the sum of the weights

of variable W_N and variable W_p. Lines 4-5 are for identifying core clusters. Line 6 is for enlarging core clusters via second-order neighbors.

Algorithm 1. Prediction of Protein Complexes
Input: The PPI network G=(V,E) PPI: protein-protein interaction network GE: gene expression data; parameter: α Output: Detected protein complexes(DP) Description: 1. Construct weighted graph W_N based on RA using Equation (5). 2. Construct weighted graph W_p based on PCC using Equation (7). 3. Construct final weighted graph W using Equation (9). 4. Forming nodes set Q according to the descending degree of node in G. 5. Take every node in Q as an initial cluster forming core complex set $C=\{C_1, C_2, \ldots, C_m\}$ by density and diameter. 6. Enlarge the core complexes based on second order neighbors and form final complex set $DP = \{DP_1, DP_2, \ldots, DP_m\}$.

(1) Density: The degree of a node V is the sum of the weights for each edge connecting to this node. Density in the weighted subgraph G=(V, E) is defined in (6). $|N|$ is the number of nodes in G and w(e) is the weight of the edge e in G.

$$m=\sum_{e=(i,j)\in E} w(e)$$

$$density(G)=\frac{2*m}{(|N|*(|N|-1))} \tag{10}$$

(2) Network Diameter: Diameter is the shortest path in a cluster.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

[1] L. Xiujuan, Y. Xiaoqin, and W. Fangxiang, "Artificial Fish Swarm Optimization Based Method to Identify Essential Proteins," *IEEE/ACM Transactions on Computational Biology & Bioinformatics,* pp. 1-1, 2018.

[2] Bo *et al.*, "Network enhancement as a general method to denoise weighted biological networks," *Nature Communications,* 2018.

[3] Z. U. Rehman, A. Idris, and A. Khan, "Multi-Dimensional Scaling based grouping of known complexes and intelligent protein complex detection," *Computational Biology & Chemistry,* vol. 74, pp. 149-156, 2018.

[4] B. Adamcsek, G. Palla, I. Farkas, I. Derenyi, and T. Vicsek, "CFinder: locating cliques and overlapping modules in biological networks," *Bioinformatics,* vol. 22, no. 8, pp. p. 1021-1023, 2006.

[5] M. Altaf-Ul-Amin, Y. Shinbo, K. Mihara, K. Kurokawa, and S. Kanaya, "Development and implementation of an algorithm for detection of protein complexes in large interaction networks," *Bmc Bioinformatics,* vol. 7, no. 1, pp. 1-13, 2006.

[6] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," (in eng), *BMC Bioinformatics,* vol. 4, p. 2, Jan 13 2003.

[7] Y. Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," (in eng), *Nature,* vol. 466, no. 7307, pp. 761-4, Aug 5 2010.

[8] V. Arnau, S. Mars, and I. Marin, "Iterative cluster analysis of protein interaction data," (in eng), *Bioinformatics,* vol. 21, no. 3, pp. 364-78, Feb 1 2005.

[9] F. BJ and D. D., "Clustering by passing messages between data points," *Science (New York, N.Y.),* vol. 315, no. 5814, pp. 972-976, 2007.

[10] A. D. King, N. Przulj, and I. Jurisica, "Protein complex prediction via cost-based clustering," (in eng), *Bioinformatics,* vol. 20, no. 17, pp. 3013-20, Nov 22 2004.

[11] A. J. Enright, S. Van Dongen, and C. A. Ouzounis, "An efficient algorithm for large-scale detection of protein families," (in eng), *Nucleic Acids Res,* vol. 30, no. 7, pp. 1575-84, Apr 1 2002.

[12] J. B. Pereira-Leal, A. J. Enright, and C. A. Ouzounis, "Detection of functional modules from protein interaction networks," (in English), *Proteins-Structure Function and Genetics,* vol. 54, no. 1, pp. 49-57, Jan 1 2004.

[13] Y. R. Cho, W. Hwang, M. Ramanathan, and A. Zhang, "Semantic integration to identify overlapping functional modules in protein interaction networks," (in eng), *BMC Bioinformatics,* vol. 8, no. 1, pp. 1-13, 2007.

[14] W. Hwang, Y.-R. Cho, A. Zhang, and M. Ramanathan, "CASCADE: a novel quasi all paths-based network analysis algorithm for clustering biological interactions," *BMC Bioinformatics,* vol. 9, no. 1, p.: 64., 2008.

[15] K. Inoue, W. Li, and H. Kurata, "Diffusion model based spectral clustering for protein-protein interaction networks," 2010.

[16] P. Lecca and A. Re, "Detecting modules in biological networks by edge weight clustering and entropy significance," *Frontiers in Genetics,* vol. 6, p. 265, 2015.

[17] T. Nepusz, H. Yu, and A. Paccanaro, "Detecting overlapping protein complexes in protein-protein interaction networks," *Nature Methods,* vol. 9, no. 5, pp. 471-472, 2012.

[18] M. Wu, X. L. Li, C. K. Kwoh, and S. K. Ng, "A core-attachment based method to detect protein complexes in PPI networks," (in English), *BMC Bioinformatics,* vol. 10(1), pp. 1-16, Jun 2 2009.
[19] F. Yu, Z. Yang, X. Hu, Y. Sun, H. Lin, and J. Wang, "Protein complex detection in PPI networks based on data integration and supervised learning method," *BMC Bioinformatics,* vol. 16, no. 12, pp. 1-9, 2015.
[20] X. Liu, Z. Yang, Z. Zhou, Y. Sun, and B. Xu, "The impact of protein interaction networks' characteristics on computational complex detection methods," *Journal of Theoretical Biology,* vol. 439, pp. 141-151, 2018.
[21] A. M. A. Maddi and C. Eslahchi, "Discovering overlapped protein complexes from weighted PPI networks by removing inter-module hubs," *entific Reports,* vol. 7, no. 1, p. 3247, 2017.
[22] B. Andreopoulos, C. Winter, D. Labudde, and M. Schroeder, "Triangle network motifs predict complexes by complementing high-error interactomes with structural information," *BMC bioinformatics,* vol. 10, no. 1, pp. 407-413, 2009.
[23] F. J, J. R, and J. T., "A max-flow-based approach to the identification of protein complexes using protein interaction and microarray data," *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM,* vol. 8, no. 3, pp. 621-634, 2011.
[24] A. Lakizadeh, S. Jalili, and S. A. Marashi, "PCD-GED: Protein complex detection considering PPI dynamics based on time series gene expression data," *Journal of Theoretical Biology,* vol. 378, no. 1, pp. 31-38, 2015.
[25] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," vol. 71, no. 4, pp. 623-630.
[26] T. Zhou, L. Lü, and Y. C. Zhang, "Predicting missing links via local information," *European Physical Journal B,* vol. 71, no. 4, pp. 623-630, 2009.
[27] C. von Mering *et al.*, "Comparative assessment of large-scale data sets of protein-protein interactions," (in English), *Nature,* vol. 417, no. 6887, pp. 399-403, May 23 2002.
[28] C. J. Wolfe, I. S. Kohane, and A. J. Butte, "Systematic survey reveals general applicability of \"guilt-by-association\" within gene coexpression networks," vol. 6, no. 1, pp. 227-0.
[29] S. Y. Pu, J. Wong, B. Turner, E. Cho, and S. J. Wodak, "Up-to-date catalogues of yeast protein complexes," (in English), *Nucleic Acids Research,* vol. 37, no. 3, pp. 825-831, Feb 2009.
[30] H. W. Mewes *et al.*, "MIPS: analysis and annotation of genome information in 2007," (in English), *Nucleic Acids Research,* vol. 36, pp. D196-D201, Jan 2008.
[31] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S. M. Kim, and D. Eisenberg, "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions," (in eng), *Nucleic Acids Res,* vol. 30, no. 1, pp. 303-5, Jan 1 2002.
[32] Tu and B. P., "Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes," *Science,* vol. 310, no. 5751, pp. 1152-1158.
[33] X. L. Li, M. Wu, C. K. Kwoh, and S. K. Ng, "Computational approaches for detecting protein complexes from protein interaction networks: a survey," (in English), *Bmc Genomics,* vol. 11, no. 1, p. S3, 2010.
[34] B. Adamcsek, G. Palla, I. J. Farkas, D. I, and T. Vicsek, "CFinder: locating cliques and overlapping modules in biological networks," (in English), *Bioinformatics,* vol. 22, no. 8, pp. 1021-1023, Apr 15 2006.
[35] S. van Dongen, "graph clustering by flow simulation," *PhD Thesis,University of Utrecht,utrecht,The Netherlands,* 2000.
[36] N. Zaki, D. Efimov, and J. Berengueres, "Protein complex detection using interaction reliability assessment and weighted clustering coefficient," *Bmc Bioinformatics,* vol. 14, no. 1, pp. 163-163, 2013.

Acknowledgements

We thanked the authors whose works were consulted and anonymous referees for constructive comments on earlier versions of the manuscript.

Funding

This research is supported by the Liaoning Natural Science Foundation Project of China (20180550918) .

Contributions

Yang Yu writes, reviews and edits the paper. Dezhou Kong writes part draft and prepares figures 6-7. All authors review the manuscript.

Corresponding author

Correspondence to Yang Yu.

Figures

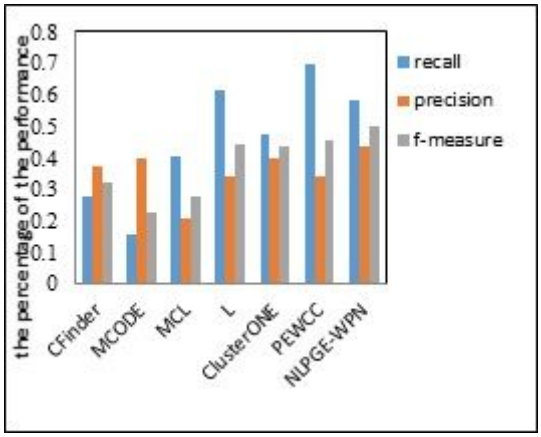

Figure 1

Comparison On Dataset CYC2008

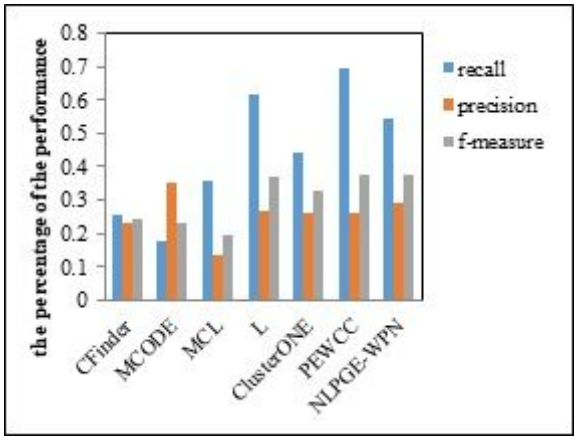

Figure 2

Comparison On Dataset MIPS

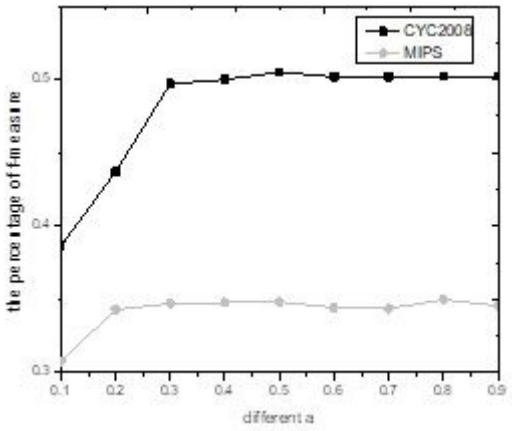

Figure 3

The influence of key parameters a on experimental results of f-measure

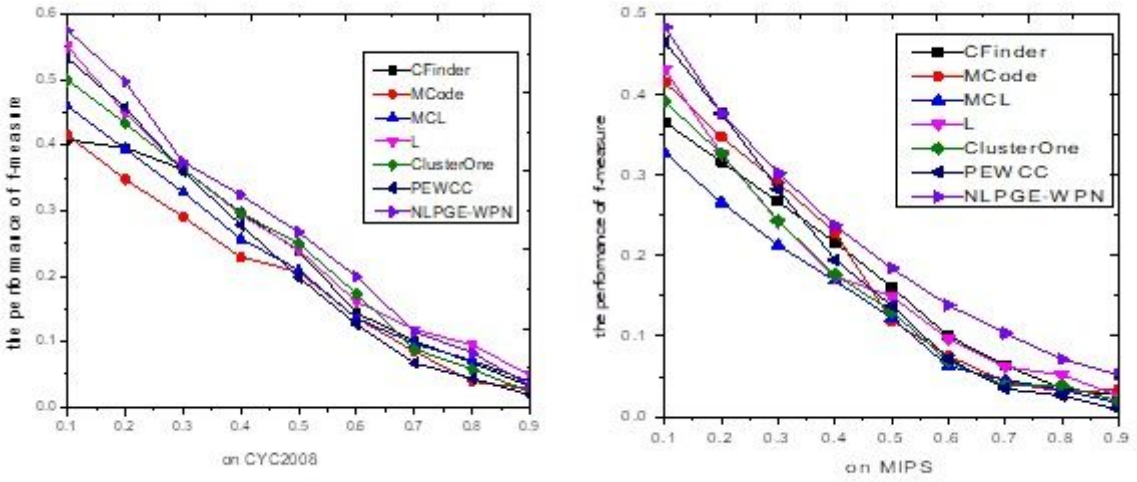

Figure 4

(a) F-measure comparisons on CYC2008. (b) F-measure comparisons on MIPS

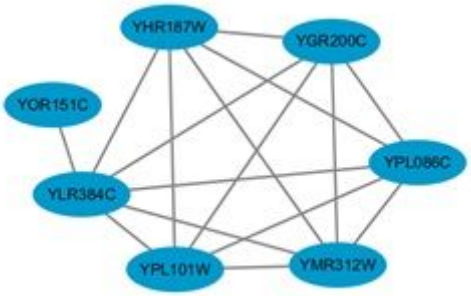

Figure 5

One example of predicted complex

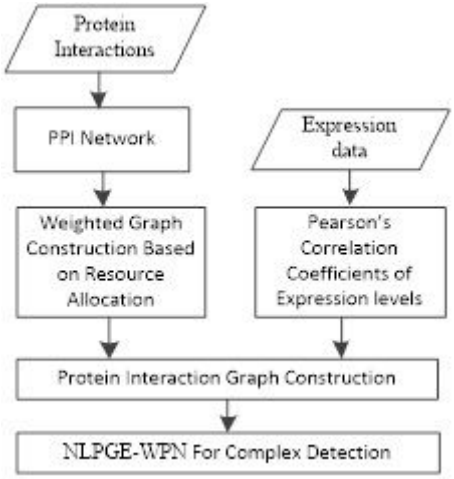

Figure 6

The workflow of our method

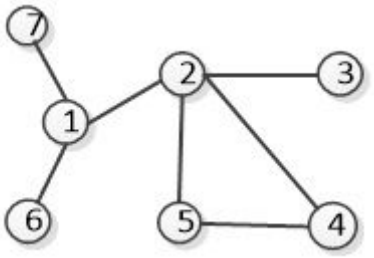

Figure 7

Sample network