Protein function module of deep learning and PPI network prediction

TaoLu¹, Sicong Huo²
¹ Nanning University, Nanning, Guangxi, 530200, China
² Nanning University, Nanning, Guangxi, 530200, China
*Sicong Huo e-mail: 237290097@qq.com

Abstract. The outbreak of new coronavirus in 2020, the development of bioinformatics technology has become a new hot spot, in which the prediction of protein function module has become a research hotspot. The existing algorithms for predicting protein functional modules have the disadvantages of low accuracy and few functional modules. In this paper, an algorithm for deep learning and PPI network prediction of protein functional modules is proposed. Firstly, three important attributes of PPI network (node location, PPI network structure and core node) are integrated. Then the improved density algorithm is used for clustering analysis, and then the improved isometric mapping dimension reduction algorithm is used for principal component analysis. Finally, MLP is used for training. The protein function module was obtained. The accuracy, F value, dimensionality reduction rate and the number of protein functional modules were compared among the other three PPI network based algorithms acc-fmd, MCL and mcode. The experimental results show that IIMP algorithm is better than other algorithms in accuracy, F value, dimension reduction rate and number of recognition modules.

1. Introduction
In recent years, with the outbreak of new coronavirus in 2020, the development of bioinformatics technology has become a new hot spot. The prediction of protein function module has become a research hotspot. The early prediction of protein functional module is based on the principle of protein sequence similarity to predict unknown protein function module. The similarity of protein sequence is detected by using blast (basic local alignment search tool) [1] and PSI-BLAST (position specific iterated blast) [2] and other tools to predict protein function module. If the sequence similarity is high, it is considered to be similar to the known protein function module, and if the similarity is low, it is considered that it is not similar to the known protein function module. However, in recent years, studies have shown that it is not rigorous to predict protein functional modules by protein sequence similarity [3], and protein functional module sequences with low similarity will also have the same functional modules. The use of PPI (protein protein interaction, PPI) network to predict protein functional modules has been recognized by many researchers, and a large number of protein functional modules have been mined. As the data of protein three-dimensional database become more and more abundant, fatcat (functional and tracer connectivity analysis toolbox) [4] and past (polypeptide angle suffix tree) [5] have become new hot spots for researchers to predict protein functional modules. PPI can be used to predict protein functional modules effectively. However, it is still a problem to predict...
proteins only by considering three-dimensional properties of proteins. Researchers have achieved good results by integrating intelligent algorithms into protein functional modules. At present, the new clustering algorithm has strong non-linear fitting ability. The linear nodes of PPI network are fitted into nonlinear nodes, which can map any complex nonlinear relationship. It can effectively predict protein functional modules. Generally, if only considering the three-dimensional properties of proteins or using intelligent algorithms, there are still some shortcomings, such as accuracy, F value is not high, and predicted protein function modules are few.

In view of these shortcomings, the reliability of protein function module prediction based on PPI network is high. The algorithm uses the three-dimensional characteristics of PPI network to predict protein functional modules (core node, node location, PPI network structure) integrate important information Methods (IDBM) was used for clustering analysis; the improved isomap algorithm (IIA) principal component analysis, and then the multi-layer perceptron (MLP) was used to train the protein function module prediction method (IIMP). Finally, the experimental results show that the accuracy and F value of IIMP algorithm are higher than those of acc-fmd, MCL and mcode.

2. Problems and definitions

PPI network is usually represented as an undirected graph G (V, e) [7], where V = {V1, V2, V3, VN}, expressed as a set of nodes, e = {E1, E2, E3 EM} is represented as the set of connected edges of each node. Node VI (1,2,3, m) is the PPI network node, and EM is the interaction between the PPI network nodes. In this paper, the network structure, node positions and core nodes in PPI are regarded as important factors to predict protein functional modules. The matrix VN ×m is represented as graph G, where each node Vij (i = 1,2,3,4 n, j=1,2,3, ⋯⋯ m) The values of are defined as follows:

\[
v_{ij} = \begin{cases} 1, & e_{ij} \in E \\ 0, & e_{ij} \notin E \end{cases}
\]

Matrix qnxm is used to represent the attributes of PPI network interaction nodes. It is defined as follows:

\[
p_{ij} = \begin{cases} 1, & n \text{ and } m \text{ Interaction} \\ 0, & n \text{ and } m \text{ not Interaction} \end{cases}
\]

By combining VN × m and QN × m, the characteristic matrix xn × m of protein is obtained, where Xi is the eigenvector of VI of PPI network interaction sample, Xij (i = 1,2, ⋯⋯, n, j=1, 2, ⋯⋯ m) The values of are defined as follows:

\[
x_{ij} = \begin{cases} V_{ij}, & j \in \{1,2,\cdots,n\} \\ Q_{ij}, & j \in \{1,2,\cdots,m\} \end{cases}
\]

The matrix yn × Z is expressed as the annotation information label of protein database which has been successfully identified, where Z is the known protein functional module category, and N is the comparison label. It is defined as follows:

\[
y_{ij} = \begin{cases} 1, & \text{Not annotated} \\ 0, & \text{Annotated} \end{cases}
\]

According to the above definition, the problem of predicting protein function module is transformed into the problem of multi label binary classification, and the mapping function H: X \rightarrow Y is obtained.

3. Protein function module of deep learning and PPI network prediction

3.1 Extract the main attributes of PPI network
According to the characteristics of PPI network, the closer the two core nodes are, the more likely it is to predict the protein function module. Through the characteristics of PPI network, the protein function module can be predicted effectively. This method has strong robustness. The noise of this method is small and robust.

Density clustering algorithm can effectively predict protein functional modules, but the traditional density clustering algorithm needs to determine the core nodes artificially, which will lead to false positive phenomenon. In this paper, a fast local optimal algorithm based on density is proposed to cluster PPI network to obtain the clusters of protein functional modules. Density fast search local optimal algorithm starts from local maximum density data every time clustering. Firstly, the local density of PPI network is calculated, and the local maximum density is used as the first large core node to cluster to form the first cluster; then the maximum density is found out from the remaining PPI network nodes which are not clustered, which will be regarded as the maximum density core node of the next cluster for clustering. The above process is repeated, and finally a threshold condition is set. When the partial density reaches the limited value, the clustering will not continue, and the sparse PPI network nodes will be treated as noise.

![Fig. 1 clustering steps of IDBM algorithm](image)

In this paper, improved density clustering algorithm is proposed to predict protein functional modules. As shown in Fig. 1 (a), the local density is calculated according to formula (5) and formula (6), and the data 7 is regarded as the first central node, and the node data with the same attribute as the central node is divided. The node data 1, 2, 3, 4, 5, 6 and 7 are classified into one category, in which 7 is the core node position of PPI network

\[ \rho_i = \sum_j x(d_{ij} - d_c) \]  

(5)

\[ \delta_i = \min_{j \in \mathcal{P}_i, \rho_j > \rho_i} (d_{ij} - d_c) \]  

(6)

The results of the first clustering are shown in Fig. 1 (b). The remaining undivided node data are 8, 9, 10, 11, 12, 13, 14, 15 and 16. Calculate their local density again to get the maximum local density data 5, in which 10 is regarded as the next PPI network core node, and then cluster analysis is carried out. The clustering result is shown in Fig. 1 (c). Check whether there are any other PPI network nodes. If not, it will end. If there are, continue to cluster. Finally, a threshold condition is set to determine whether the remaining data nodes are regarded as noise. As shown in Fig. 1 (d), when the local density of the whole PPI network reaches the limited density value, clustering will not continue, and sparse PPI network nodes will be treated as noise.
3.2 principal component analysis

As mentioned above, the three main attributes (core node, core node location, PPI network structure) in PPI network play a key role in protein function module prediction. These three main attributes are used to predict protein function module to improve accuracy and reduce data noise. In this paper, multi-layer perceptron (MLP) is used in training. If the cluster is trained, the dimension of cluster is too high, which leads to high noise and affects the prediction result. Considering the correlation of protein functional attributes, it is necessary to conduct principal component analysis on the clusters after clustering. This paper proposes an improved isometric mapping dimensionality reduction algorithm, which can maintain the consistency of data structure before and after dimensionality reduction. The basic idea is as follows: use the cluster after clustering to establish a matrix; then calculate the minimum distance between any cluster nodes, expressed as geodesic distance; finally, use the (multiple dimensional scaling, MDS) [8] algorithm to get low dimensional data.

The process of improving isometric mapping dimensionality reduction algorithm is as follows: (1) construct adjacency matrix. Cluster is regarded as neighborhood point, ε is neighborhood radius, and LJ is neighborhood point of Li when \( |Li-LJ| \leq ε \). The formula is as follows:

\[
W_{ij} = \begin{cases} 
1: \text{L_i and L_j, Domain point} \\
0: \text{Other information}
\end{cases}
\]  

(7)

(2) The shortest path is calculated. If Li and LJ are domain points, the distance between these two points is replaced by the Euclidean distance \( s(I, J) \)

\[
S_i(L_i, L_j) = \min \{ S_i(L_i, L_j), S_i(L_k, L_j) + S_i(L_k, L_j) \} \]  

(8)

The shortest path moment of geodesic distance between two points can be obtained by formula (7) (8). SC = [SC (I, J)] n x m, then w matrix is applied to MDS algorithm to calculate the dimension reduction. In order to improve the quality of dimension reduction, the geodesic line is modified as follows, and formula (9) is used for calculation:

\[
S_c = \left[ s_c(x_i, x_j) \right]_{n \times N}, i, j = 1, 2, \ldots, N
\]  

(9)

Where SC (Xi, XJ) is determined by the following formula (10):

\[
x_c(x_i, x_j) = \min \left\{ \sum_{k=1}^{N} s(x_k, x_{\omega k}) \right\} + s(x_i, x_j) + \sum_{k=1}^{N} d(x_{\omega k}, x_{\omega k}) \]  

(10)

The distance \( s \) shown in formula (11) is Euclidean distance, and the kernel matrix K is calculated by geodesic distance:

\[
K = HS\text{T}H
\]  

(11)

Where: the central matrix H is calculated by the following formula:

\[
H = I_n - \frac{1}{N} e_n e_n^T
\]  

(12)

Where \( e_n = [0 \ 1]^T \in \mathbb{R}^N \). We decompose the matrix K to obtain the largest n eigenvalues \( \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n \), and the corresponding eigenvector \( \omega_1, \omega_2, \omega_3, \ldots, \omega_n \) by using the matrix \( \omega = \{ \omega_1, \omega_2, \omega_3, \ldots, \omega_n \} \) obtains the corresponding low dimensional data \( x_i \):

\[
X_i = \left\{ x_i \mid x_i \in \mathbb{R}^S, j = 1, 2, \ldots, N \right\}
\]  

(13)

3.3 multilayer sensor training and functional protein selection
The core cluster of PPI network after clustering and dimensionality reduction reflects the importance of protein functional modules in PPI network. When the importance of core cluster nodes is higher, the generation represents the more functional attributes of proteins. In this paper, the reduced core cluster is regarded as a feature matrix and combined with the matrix formed by the known principal components of protein functional module attributes to obtain a new feature matrix. In order to improve the quality of training, min max normalization is applied to the feature matrix to make each element data linear transformation.

In this paper, we combine the reduced core cluster elements in PPI network with the principal components of known protein functional module attributes to obtain a feature matrix $F_{IJ}$, where the elements $f_i = (F_{I1}, f_{i2}, \ldots)$ The value of $F_{IJ}$ is shown in formula (14)

$$f_{ij} = \begin{cases} Y_{ij}, & j \in \{1,2,\cdots,n\} \\ X_{ij}, & j \in \{1,2,\cdots,n_z\} \end{cases}$$

(14)

Matrix $x'_{ij}$ is represented as clusters in PPI network after dimensionality reduction, and $Y_{ij}$ is represented as features of attribute principal components of protein functional modules. The principal component of protein function module attribute in PPI network reflects the functional information of protein, and the elements of core cluster of PPI network reflect the importance of PPI network in which the protein function module is located. The higher the importance of core node of cluster, the more functions the protein has [9].

In order to improve the convergence speed of IIMP algorithm, min max normalization is used to obtain the normalized matrix $f'_{ij}$. The linear transformation result of each element data falls into the interval $[0,1]$, as shown in formula (15). $f_i=(f_{i1}, f_{i2}, \ldots, f_{in})$ denotes the attributes of protein functional modules, $F_J (f'_i, f'_{i1}, f'_{in})$ is expressed as a normalized eigenvector. As shown in formula (15).

$$f'_{ij} = \frac{f_{ij} - \min(f_{ij})}{\max(f_{ij}) - \min(f_{ij})}$$

(15)

In this paper, a single hidden layer multi-layer perceptron (MLP) is used as the classifier. The activation function and loss function obtained by clustering PPI network are modeled, and the complex relationship between predicted protein functional modules and known protein functional modules is mined by using the hidden layer of MLP perceptron. As shown in Figure 2, this paper takes the feature vector after the term of known protein function module is characterized as input, and the annotation of known protein function module is taken as output to establish the mapping relationship between multi feature and multi-function in protein function module.

![Fig. 2 Schematic diagram of multilayer sensor](image-url)
The number of nodes in the input layer \( x \) is the feature vector of the cluster obtained by isometric mapping dimensionality reduction algorithm. The hidden layer outputs \( f(w1x + B1) \), where \( W1 \) is the weight (also known as the connection coefficient), \( B1 \) is the offset, and the number of nodes in the output layer is the number of notes of the predicted protein function module in the data set.

The number of hidden layer nodes is according to the empirical formula \([10]\), the empirical formula is \( n \) dimensional space straight line \( y = a_1x_1 + a_2x_2 + ... + a_nx_n \), where \( a_1, a_2 a_n \) is an empirical constant. The hidden layer uses the relu activation function, such as formula (16).

\[
RReLU(x) = \max([-\infty, x) \cup (x, +\infty)]
\]

(16)

Tanh activation function is used in the output layer, such as formula (17)

\[
tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}
\]

(17)

The cross entropy is used as the loss function of the output layer, such as formula (18)

\[
H(p, q) = -\sum_x (p(x) \log q(x) + (1 - p(x)) \log(1 - q(x)))
\]

(18)

The neural network is trained by batch learning \([11]\), the batch size is 20% of the protein number in the training set, the iteration number is 800, the learning rate is 0.08, and the momentum is 0.9.

4. experimental environment and data set

The environment used in this experiment is Windows 7 operating system, Intel dual core processor, physical memory 8g, processor speed is 3.1GHz. The programming tool is r3.32. CPU acceleration technology and mxnet deep learning framework are used to train MLP.

In order to verify the effectiveness of the IIMP algorithm proposed in this paper, yeast PPI interaction network data was selected as experimental data. The specific experimental data, using yeast PPI interaction network data, dip \([12]\) database (version dip20200719), cogs database providing gene ontology, Interpro provides a database of domains and functional sites in proteins Using uniprotkb / Swiss prot for ID conversion of known protein functional modules Then, the effective protein terms were screened by tandem affinity purification technology.

4.1 evaluation index

In this paper, precision, recall and F-measure is the standard to judge the advantages and disadvantages of the algorithm. \([13]\) The formula is as follows:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

(19)

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

(20)

\[
\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

(21)

For the selection of protein function module, the matching degree between the function module a mined by the algorithm and the known function module \( B \) is calculated by formula (22). If the matching degree between the identified function module \( a \) and the known function module \( B \) exceeds the given threshold value, the known function module is said to be identified. According to the threshold value set in reference \([14]\), 0.2 [2160-2162] will be set. If \( OS (a, b) \gg 0.2 \) is to predict that protein functional modules will be recognized by known protein functional modules.

\[
OS(A, B) = \frac{|A \cap B|}{|A||B|}
\]

(22)

4.2 analysis of experimental data
In order to analyze IIMP algorithm to detect protein function modules on PPI network in detail, the IIMP algorithm was compared with acc-fmd $^{[15]}$, MCL $^{[16]}$, mcode $^{[17]}$ algorithm. Acc-fmd has strong flexibility and strong robustness, and can obtain high-quality data in the face of similarity problems; MCL algorithm can easily obtain data with good training quality, and the accuracy of the data obtained after clustering Mcode algorithm can easily find the close relationship between nodes in PPI network, and it will not bring high false positive effect due to high-throughput data, and the obtained data has high accuracy. In the experiment, the IIMP algorithm is compared with each algorithm, and the number of function modules extracted from PPI network provided by dip database is analyzed.

As shown in Figure 3, the number of protein functional modules obtained by DBSCAN is 623, and the number of protein functional modules obtained by the improved density clustering algorithm IDBM is 925, and the number of clustering modules is significantly increased.

As shown in Fig. 4, Isomap dimensionality reduction rate of Isomap is 50.2%. The dimension reduction rate of protein functional modules obtained by the improved isometric mapping algorithm MDS Isomap is 72.1%, and the dimension reduction rate of the improved algorithm is significantly increased.
As shown in Figure 5, PM represents the total number of protein functional modules mined by each algorithm, full refers to the number of fully identified protein functional modules, and TP refers to the number of matching with known protein functional modules. The total number of protein function modules mined by IIMP algorithm is 392, among which 253 are matched. The number of matched protein function modules is the largest among the four algorithms, which indicates that the IIMP algorithm proposed in this paper is more efficient.

Fig. 6 the number of clusters before and after dimensionality reduction of various algorithms
As shown in Fig. 6 and Fig. 7, in PPI network, the dimension reduction rate of protein attribute principal component feature extraction results obtained by different clustering algorithms is represented, where M represents the cluster of PPI network obtained after clustering, n represents the cluster obtained through dimensionality reduction, 1-N / M represents the dimensionality reduction rate, and the dimensionality reduction rate of IIMP algorithm is 75.36%, which are higher than acc-fmd, MCL and mcode algorithms. The dimensionality reduction rates of other algorithms are 40.70%, 54.71% and 43.96%, respectively. The IIMP algorithm has the highest dimensionality reduction rate among the four algorithms.

Fig. 8 comparison of significant protein modules of each algorithm
Fig. 9 significant module number ratio of each algorithm

As shown in FIG. 8 and FIG. 9, SC represents the significant number of protein functional modules. In the number of functional modules mined by IIMP algorithm, the proportion of significant modules reached 39.03%. Compared with acc-fmd, MCL and mcode, the proportion of other three algorithms was 22.73%, 28.57% and 38.58%. Thus, IIMP algorithm mining function module is more biological statistical significance.

Fig. 10 efficiency analysis of IIMP algorithm and other three algorithms

As shown in Figure 10, this paper uses the leave one out method to predict the IIMP algorithm function module. Proteins have been annotated in MF, BP and CC networks. When predicting protein function modules, all the clustering data are divided into k small data sets, and then k-1 data sets are used as training data sets, and the remaining one is used as test data sets. Although this method is time-consuming, the overall results are close to the expected values. In the experiment, the precision, recall rate and F value of PPI network provided by dip database are compared. In terms of accuracy, the accuracy of IIMP algorithm is 0.381, which is higher than acc-fmd algorithm 0.231, MCL algorithm 0.129, mcode algorithm 0.328; in terms of recall rate, IIMP algorithm is higher than acc-fmd, mcode algorithm, only lower than MCL algorithm. From the F value, IIMP algorithm is higher than acc-fmd, MCL, mcode algorithm, has obvious advantages. In general, IIMP algorithm can effectively predict protein functional modules, and is better than acc-fmd, MCL and mcode algorithms in accuracy, F value, number of recognition modules and dimensionality reduction rate.
5 Conclusion
With the continuous growth of PPI network data, how to predict protein functional modules efficiently is an important research problem in modern bioinformatics. The prediction of protein function module based on PPI network is a popular prediction method. This kind of method is low cost and efficient, but the obtained data has a lot of noise. This paper proposes a IIMP algorithm to solve this problem. This algorithm integrates the three core information (PPI network structure, core nodes, and positions of each node) in PPI network, and comprehensively considers various elements in PPI network. In this paper, three methods and technologies based on density clustering algorithm, isometric mapping dimensionality reduction algorithm and multi-layer perceptron are used to predict protein functional modules. In order to verify that IIMP algorithm is more effective than other algorithms, PPI network data and known protein function module terms are downloaded from dip database, corresponding protein function module numbers are obtained from interpro database, and cogs function annotation is used. The experimental results show that IIMP algorithm can effectively predict protein functional modules, and is better than acc-fmd, MCL and mcode algorithms in accuracy, F value, dimension reduction rate and number of recognition function modules.

In the future work, we can further study and predict protein functional modules from the following directions. 1) In depth understanding of the impact of PPI network density on the prediction of protein function module, in different PPI network density affects the size of clustering, data preprocessing should be how to divide the density of PPI network, different density for the prediction of protein function module is great, this problem needs to be further studied; 2) based on the understanding of PPI network, it is necessary to study How to use modern popular intelligent optimization algorithm such as Bayesian intelligent algorithm, neural network intelligent algorithm to predict protein function module; 3) PPI network is a dynamic network, most researchers are taking static PPI network for research, how to predict protein function module and PPI network dynamic when the overall structure of PPI network has not changed It is also a new research direction to predict different protein functional modules under the condition of.

Acknowledgment
Scientific Research Foundation Enhancement Project for young and middle-aged Teachers in Guangxi Universities in 2020
Agricultural products logistics loss monitoring and control system based on big data visualization technology
Project No.: 2020KY64015
The 2020 Professor Training Project of Nanning University
Application of image enhancement algorithm based on Gamma correction in temperature monitoring system
Project No.: 2020JSGC03

References
[1] Lin min, Zhong Yiwen, Lin Juan. Local alignment algorithm of biological sequences based on GPU parallelism [J]. Journal of Fujian agricultural and Forestry University (NATURAL SCIENCE EDITION), 2015,44 (04): 442-448Dynamic weighted PPI network complex mining based on ant colony clustering [J / OL]. Computer application research: 1-10 [2019-06-13]. Https://doi.org/10.19734/j.issn.1001-3695.2018.07.0518.
[2] Chen B, Fan W, Liu J, et al. Identifying protein coMLPexes and functional modules--from static PPI networks to dynamic PPI networks.[J]. Briefings in Bioinformatics, 2014, 15(2):177.
[3] Liu Wei, Liu Shang, Zhou Xuan. Study on BP neural network sub batch learning method [ J ]. Journal of intelligent systems, 2016, 11 (2): 226-232.
[4] Liu Hongxin. Research on PPI network functional module detection algorithm based on ant colony clustering mechanism [D]. Beijing University of technology, 2013
[5] Hu Qingsheng, Lei Xiujuan. Improved Markov clustering algorithm for PPI network [J]. Computer science, 2015, 42 (7): 108-113

[6] Lei X, Li H, Zhang A, et al. iOPTICS-GSO for identifying protein coMLPexes from dynamic PPI networks[J]. Bmc Medical Genomics, 2017, 10(Suppl 5):80

[7] Zhang X F, Ouyang L, Zhu Y, et al. Determining minimum set of driver nodes in protein-protein interaction networks[J]. Bmc Bioinformatics, 2015, 16(1):146.

[8] Song J, Singh M. How and When should Interactome-derived Clusters be Used to Predict Functional Modules and Protein Function[J]. Bioinformatics, 2009, 25(29):3143-3150.

[9] YE Y, GODZIK A. FATCAT: a Web server for flexible structure comparison and structure similarity searching [J]. Nucleic Acids Research, 2004, 32( Web Server issue) :W582 -W585.

[10] TUBIG H, BUCHNER A, GRIEBSCH J. PAST: fast structurebased searching in the PDB [J]. Nucleic Acids Research, 2006,34( Web Server issue) : W20 - W23.

[11] Zhang X F, Ouyang L, Zhu Y, et al. Determining minimum set of driver nodes in protein-protein interaction networks[J]. Bmc Bioinformatics, 2015, 16(1):146

[12] LIU W, LIU S, ZHOU X. Subbatch learning method for BP neural networks [ J ] . CAAI Transactions on Intelligent Systems, 2016, 11( 2) : 226 – 232.

[13] Song J, Singh M. How and When should Interactome derived Clusters be Used to Predict Functional Modules and Protein Function[J]. Bioinformatics, 2009, 25(29):3143-3150

[14] LIU W, LIU S, ZHOU X. Subbatch learning method for BP neural networks [ J ] . CAAI Transactions on Intelligent Systems, 2016, 11( 2) : 226 – 232

[15] Liu Hongxin. Research on PPI network function module detection algorithm based on ant colony clustering mechanism [D]. Beijing University of technology, 2013

[16] Hu Qingsheng, Lei Xiujuan. Improved Markov clustering algorithm for PPI networks [J]. Computer science, 2015, 42 (7): 108-113

[17] Lei X, Li H, Zhang A, et al. iOPTICS-GSO for identifying protein coMLPexes from dynamic PPI networks[J]. Bmc Medical Genomics, 2017, 10(Suppl 5):80