

“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# Detecting Overlapping Protein Complexes in Dynamic Protein-Protein Interaction Networks by Developing a Fuzzy Clustering Algorithm

Ruiping Yin<sup>1,2</sup>, Kan Li<sup>1</sup>

<sup>1</sup>. School of Computer Science & Technology  
Beijing Institute of Technology  
Beijing, China  
{yrp, likan}@bit.edu.cn

Guangquan Zhang<sup>2</sup>, Jie Lu<sup>2</sup>

<sup>2</sup>. Center for Artificial Intelligence, Faculty of  
Engineering and Information Technology  
University of Technology Sydney  
Sydney, Australia  
{Guangquan.Zhang, Jie.Lu}@uts.edu.au

**Abstract**—Protein complexes play important roles in protein-protein interaction networks. Recent studies reveal that many proteins have multiple functions and belong to more than one different complexes. To get better complex division, we need to consider time-dependent information of networks. However, only few studies can be found to concentrate on detecting overlapping clusters in time-dependent networks. To solve this problem, we propose integrated model of time-dependent network (IM-TDN) to describe time-dependent networks. On the base of this model, we propose similarity based dynamic fuzzy clustering (SDFC) algorithm to detect overlapping clusters. We apply the algorithm to synthetic data and real world protein-protein interaction network dataset. The results showed that our algorithm by using the model which we proposed achieved better results over the state-of-the-art baseline algorithms.

**Keywords**—fuzzy clustering; protein-protein network; temporal networks

## I. INTRODUCTION

Protein interactions are the physical contacts between two or more protein molecules. They are crucial components of all cellular processes. The interactions can be described as protein-protein interaction (PPI) networks [1], where nodes represent proteins and links represent protein interactions. In PPI networks, proteins bind with other proteins to perform biological functions as the form of protein complexes [2] [3]. Complex detection of PPI networks is a clustering problem that divides nodes into different groups, in which the nodes connect each other more tightly than those out of the groups [4] [5]. The task can be used to understand the behavior of organisms, predict the function of proteins and treat the disease.

This work faces a problem that many proteins have multiple functions and play important roles in different complexes. We need to detect the overlapping protein complexes in the PPI networks [6]. Meanwhile, PPI networks evolve over time. We named it dynamic PPI networks or time-dependent networks. At the different stages of interactions, the weight of interactions will change. Some proteins may disappear and new proteins may join in. So, we also need to consider the dynamic of the process.

Recently, algorithms have been proposed to solve these problems. Some of them considered the overlapping [7] of protein complexes. They treat PPI network as a static network and ignore the historical status [8]. Others take the temporal into

account [9] [10]. However, they fail to detect completed complexes. Because one protein which belongs to a complex cannot belong to another. How to utilize the temporal information in the evolving process to detect overlapping complexes of time-dependent networks is still challenging.

The main contributions of this paper are:

1) *Aiming at solving the drawbacks above to get better results of complexes division for PPI networks, we developed an integrated model to describe the dynamic of PPI networks in which different stages of time-dependent network was synthesized into a similarity matrix with different weights.*

2) *We also define similarity of nodes in the integrated model. Meanwhile, one protein may belong to different complexes. We can regard this problem as a fuzzy clustering problem. So we proposed a similarity based dynamic fuzzy clustering (SDFC) algorithm to detect overlapping complexes.*

The paper is organized as follows. In Section II, we introduce related work in this topic. Section III shows the model IM-TDN to integrate the temporal information and propose a definition of similarity of nodes in time-dependent networks to extract node similarity information. In section IV, we propose a novel SDFC algorithm to detect protein complex structure of dynamic PPI networks. Section V makes experiments on synthetic data and real world data to illustrate our algorithm and related concepts. Finally, in section VI, conclusions and future work are discussed.

## II. RELATED WORK

A PPI network is a biomolecule relationship network that plays an important role in biological activities. Studies of functional complexes in a PPI network contributed greatly to the understanding of biological mechanism. A lot of work has been done to improve the quality of the detection result.

At the beginning of these studies, they treat PPI network as a static network to detect complexes of PPI. In [11], Bader and Hogue proposed the three-stage Molecular Complex Detection algorithm to identify densely connected proteins from a PPI graph. In [12], Enright et al. proposed Markov Clustering algorithm. It is an unsupervised clustering algorithm for a graph. The algorithm simulates random walks within a graph by alternation of two operators named expansion and inflation.

Getz et al. [13] proposed the Super Paramagnetic method to group PPI networks of diverse function, inhomogeneity and different sizes. Different methods above have their own characteristics and ranges of application.

However, what they deal with is static PPI networks. They lost much information in the process of clustering. Standard clustering is not ideal for PPI network. Because nodes may belong to several complexes. Fuzzy clustering has been used to detect complexes in PPI networks. Nepuzė [7] introduced overlapping neighborhood expansion (CluserONE) to detect overlapping complexes. Cao [14] proposed a method of pseudo-clique extension based on fuzzy relation to detect protein complexes from weighted PPI network. There are also other methods to detect overlapping complexes. Wang [15] utilized hierarchical clustering algorithm to reveal the complexes in PPI networks which can identify both overlapping and hierarchical functional modules.

There are also some methods use dynamic networks to describe PPI. Literature [16] propose a novel technique to detect temporal protein complexes from the dynamic PPI networks. The author construct a series of dynamic PPI networks by detecting stable interactions and transient interactions by integrating protein interaction data and gene expression data. Based on the concept of overlapping temporal communities, the author propose a novel Time Smooth Overlapping Complex Detection model. However, in dynamic PPI networks, temporal networks evolve over time, and different temporal networks should take different influences to complex detection of PPI network. Taylor et al. [17] proposed another method to deal with dynamic PPI networks, where current complex detection of PPI networks is only related with the prior temporal network next to it. While the method can get better result of complex detection than considering single static PPI network, it discards some time information. Because in dynamic networks, past temporal networks can influence the complex detection of current PPI networks. Therefore, the key problem is how to utilize the temporal networks to detect complexes of current PPI network.

In this paper, we propose a novel model that considers time information and a new algorithm to reveal the overlapping complexes in dynamic PPI networks.

### III. TIME-DEPENDENT NETWORKS MODEL

In this section, we propose a new model to integrate the temporal information of time-dependent networks into a final matrix. We will describe time-dependent networks first. Then we discuss how to deal with time-dependent information in PPI networks according to their evolving process and define a new definition of node similarity to extract complex information. Finally, we integrate the similarity information of temporal networks together and model time-dependent networks to a final similarity matrix.

#### A. Definition of time-dependent networks

A Time-dependent network is dynamic networks changing with times [9]. As shown in Fig. 1, a time-dependent network is a serial of simplex networks. At a fixed time  $t$ , the network corresponds to a simplex network with same node sets.



Fig. 1. Schematic diagram of a time-dependent network

We can use (1) to describe a time-dependent network which contains  $T$  temporal networks.

$$N = [N_1, N_2, N_3, \dots, N_{T-1}, N_T] \quad (1)$$

#### B. Integrated model of time-dependent networks

In time-dependent networks, if we detect complexes just according to current temporal network, there will be bias because there may be false interactions in single temporal network. Meanwhile, detecting protein complexes by using a serial of networks is more reasonable. At the same time, we can also discover the evolution of the protein. Therefore, we need to consider past temporal networks, but how many temporal networks we need to consider and how to utilize these temporal networks are the key problems to be resolved.

When we analyze complexes of current temporal network, we need to consider the prior temporal network next to current temporal network, because current temporal network develops from it. More specially, what we should consider is the interactions in the prior temporal network different from the current temporal network. And the more the different interactions are, the more we should consider the last temporal network. Recurrently, the prior temporal network has a similar relation with its prior temporal. Therefore, we can get the relationship of different temporal networks we should consider as (2).

$$\begin{cases} I'_t = \alpha_1 I_t + (1 - \alpha_2) I'_{t-1} \\ I'_{t-1} = \alpha_2 I_{t-1} + (1 - \alpha_2) I'_{t-2} \\ \dots \\ I'_{t-n+1} = \alpha_n I_{t-n+1} + (1 - \alpha_n) I'_{t-n} \end{cases} \quad (2)$$

where  $I_t$  is the interaction relationship of temporal network at time  $t$ ,  $I'_t$  is the fixed relationship and  $I_0 = I'_0$ ,  $\alpha_i$  takes value between  $[0, 1]$ , and it can be got by (3).  $\alpha_i$  describes the proportion between temporal network at time  $t - i + 1$  and its prior temporal network at time  $t - i$ , when we detect complexes of dynamic networks.

$$\alpha_i = \frac{\sum E_{t-i+1}(a,b) * E_{t-i}(a,b)}{\sum E_{t-i+1}(a,b) + \sum E_{t-i+1}(a,b) - \sum E_{t-i+1}(a,b) * E_{t-i}(a,b)} \quad (3)$$

where the numerator describes the same interactions between temporal network  $t - i + 1$  and temporal network  $t - i$  and denominator represents all interactions between temporal network  $t - i + 1$  and temporal network  $t - i$ .  $E_{t-i+1}(a, b)$  and  $E_{t-i}(a, b)$  stand for respectively values between node a and b in temporal network  $t - i + 1$  and temporal network  $t - i$ .

In (2), we describe the model in an iteration formula, we can rewrite it as (4).

$$I_t = \alpha_1 I_t + (1 - \alpha_1) \alpha_2 I_{t-1} + \dots + (1 - \alpha_1) (1 - \alpha_2) \dots \alpha_n I_{t-n+1} \quad (4)$$

Now we analyze utilization of temporal information. There are two situations. One is that if  $\alpha_i$  is small and  $1 - \alpha_i$  is large, that is to say,  $I_{t-i+1}$  has a big difference with  $I_{t-i}$ . That means there is mutation between  $I_{t-i+1}$  and  $I_{t-i}$ . Under this condition, the temporal networks before time-slot  $t - i$  should not be considered and we just consider temporal networks between  $t$  and  $t - i + 1$ . The other is that when there is no mutation and the coefficient of  $I_{t-i+1}$  in (4) is small enough, the corresponding terms can be ignored.

### C. Interaction information extraction

$I_{t-i}$  in (2) and (4) describes the interaction information of nodes. If we use adjacent matrix as  $I_{t-i}$ , the links in temporal networks can only take values 0 or 1. However, the two values are not enough to describe interaction among nodes of temporal networks precisely and fully. For example, for two nodes, even though they are not connected directly by a link, it doesn't mean that there is no interaction between them, because the two nodes might be connected by a path containing more than one edge. But such interaction can't be measured by values 0 or 1. Therefore, it is necessary to excavate the interaction information of temporal networks more exactly. Here, we define a node similarity to measure the interaction information of temporal networks.

Jaccard's coefficient, Salton index, Sorenson index, Common neighbors, Adamic-Adar and Leicht-Holme-Newman (LHN) [18], [19] are five famous methods for node similarity. Zhou Tao and Liben-Nowell have proved that common neighbors are the simplest method "Common neighbors" which usually performs surprisingly well. However, "Common neighbors" requires that there are common neighbors between two nodes, otherwise the similarity of two nodes will be 0. Literature [20] resolved the problem above and defined the similarity between two nodes connected by only one edge without common neighbors. But similarity of two nodes is 0 if the two nodes are not connected by directly one edge. Here, we propose a new method to compute node similarity between node  $v_i$  and  $v_j$  described as (5).

$$\text{Sim}(v_i, v_j) = \begin{cases} 1, & \text{if } v_i \text{ and } v_j \text{ are same} \\ \frac{\sum_{B \in P} \prod_{E(u,w) \in B} S(u,w)}{\text{if there are paths between } v_i \text{ and } v_j} & (5) \\ 0, & \text{otherwise.} \end{cases}$$

$$S(u,w) = \frac{1}{\text{deg}(u) + \text{deg}(w) - 1} \quad (6)$$

In (5),  $P$  is the set of paths between node  $v_i$  and  $v_j$ ;  $E(u,w)$  is one edge belonging to the path  $B$  and  $u, w$  are two nodes of

the edge  $E(u,w)$ . In (6),  $\text{deg}(u)$  and  $\text{deg}(w)$  are degree of nodes  $u$  and  $w$  respectively.

The node similarity considers all the paths between two nodes that breaks the restriction above. Node similarity is a number that reflects the similarity of two nodes in simplex networks. Therefore, we can compare and accumulate the node similarity of different simplex networks. Besides, node similarity can help to detect complex structure because two nodes owning larger similarity can be divided into one complex.

We utilize node similarity information of temporal networks and (4) to integrate node similarity information of different temporal networks into current temporal network. In this way, we can model the time-dependent network to a similarity matrix that the similarity value between each node has been calculated. Then we propose a fuzzy clustering algorithm to detect protein complexes.

## IV. ALGORITHM

In this section, we propose the SDFC algorithm to detect protein complexes in time-dependent PPI networks. To describe the process of our algorithm, we introduce the concept of clustering at first. Then, we give the details of our algorithm in Part B.

### A. Concept of clustering algorithm

Clustering is an important tool in complex network analysis, which are defined intuitively as the nodes in the same group is more tightly to each other than the nodes between groups [21]. Given the network adjacency matrix  $N$ , where the element  $N_{ij}$  is the connection strength between node  $i$  and node  $j$ . Hard clustering is to construct a partition  $C = (C_1, C_2, \dots, C_k)$ . In formula, we can represent this partition as a matrix  $U_{n \times k}$ , where  $n$  is the number of nodes.  $U_{ij}$  is 1 when node  $i$  belong to  $C_j$  and 0 otherwise. Each node belongs to a unique cluster  $C_i$ .

Fuzzy clustering is different from hard clustering. It can be used to detect clusters by using grade of membership [21]. It attempts to partition a finite collection of  $n$  element  $X = \{x_1, x_2, \dots, x_n\}$  into a collection of  $k$  clusters with respect to some given criterion. Let  $U$  be the fuzzy membership matrix, where  $u_{ij} \in [0,1]$ ,  $1 \leq i \leq k$ ,  $1 \leq j \leq N$  is the grade of membership of data  $x_k$  to the cluster and  $\sum_{i=1}^k u_{ij} = 1$ . In fuzzy clustering, each data point can belong to different clusters which corresponds to the multiple functions of proteins in different complexes.

### B. Similarity based dynamic fuzzy clustering algorithm

Fuzzy C-means clustering (FCM) is the most widely used algorithm in fuzzy clustering. However, the input of FCM is a set of data points. And it requires a function to compute the distance between nodes and central of clusters. Thus, we cannot use this algorithm directly. Our dataset is the adjacent matrix of time-dependent PPI networks. We cannot compute the central of clusters directly because that the adjacent matrix only represents the distance between nodes. Meanwhile, we also need this algorithm to utilize the temporal information.

Input	One time-dependent PPI network $N$ containing $m$ time slots.
Output	$Complexes = \{C^{(1)}, C^{(2)}, \dots, C^{(s)}\}$ , where $C^{(s)}$ is the $s$ th complex.
Step1	Compute similarity of each node according to (5)
Step2	Integrate the serial information of $m$ time slots to get a final network $L$ according to (3) and (4).
Step3	Initialize the membership matrix $U_{n \times s}$ in the constraint of $\sum_{i=1}^s u_{ij} = 1$ from a Gaussian distribution.
Step4	Compute the cost function (9) and the gradient until it less than a threshold. If the cost is less than a threshold, the algorithm will terminate.
Step5	Compute the membership matrix $U$ and go to step 4.

Therefore, we propose similarity based dynamic fuzzy clustering algorithm (SDFC) which consider time  $t$  as a parameter in the objective function to integrate the temporal information in the clustering process. Our approach also considers practical aspects of the task and introduces some parameters to improve the results. We utilize parameters to control the objective function. In the algorithm, we use the model which is introduced in Section □ to construct a final matrix. The matrix  $L$  is defined as (7):

$$L = \sum_{i=1}^m \alpha_i A^{(i)} \quad (7)$$

where  $m$  is the time stamp that we need consider. When  $m$  is too small, we cannot solve the problem that the interactions between nodes mutate. If the  $m$  is too large, we will introduce much useless information. Especially, when  $m = 1$ , it is the traditional model used in static complex networks.  $A^{(i)}$  is node similarity matrix of  $i$ th time and  $\alpha_i$  is the proportion which represents the similarity between time  $i$  and current time. In order to detect clusters by using node similarity, we use (5) to calculate the similarity between each node. Then we use (3) to integrate the similarity information of different time stamps with different weight.

After constructing the final matrix, we obtain similarity matrix  $L$  which corresponds to the actual similarity of nodes in time-dependent networks. This leads to the following equation (8), which represents the fitness of the actual similarity and the partition.

$$J(U, \alpha_1, \dots, \alpha_T) = \sum_{i=1}^N \sum_{j=1}^N (\sum_{t=1}^T \alpha_t A_{ij}^{(t)} - \sum_{k=1}^C u_{ki} u_{kj})^2 \quad (8)$$

where  $u_{kj}, u_{kj} \in (0,1)$  is the grade of membership and  $\sum_{k=1}^C u_{ki} u_{kj}$  is the similarity of this partition between node  $i$  and node  $j$ ;  $\sum_{t=1}^T \alpha_t A_{ij}^{(t)}$  is the actual similarity between node  $i$  and node  $j$ ;  $\alpha_t$  is the network similarity value between time stamp  $t$  and current time;  $A_{ij}^{(t)}$  is the distance between node  $i$  and node  $j$  at time stamp  $t$ .

We must find the value of these parameters which minimize (8) while satisfying the conditions of  $\sum_{i=1}^c u_{ij} = 1$  and  $0 < \sum_{k=1}^N u_{ik} < N$ . We also need to make  $\alpha_t$  be close to 1 to ensure

the time-dependent network is soft. So, we just treat  $\alpha_t$  which is calculated in (3) as an initial value to reduce some noise in the dataset. We can use the following objective function (9) to get the minimum value of (8).

$$J(U, \alpha_1, \dots, \alpha_T, \lambda_1, \lambda_2, \dots, \lambda_{n+T}) = J(U, c_1, c_2, \dots, c_c) + \sum_{i=1}^N \lambda_i (\sum_{k=1}^C u_{ki} - 1) + \sum_{t=1}^T \lambda_{n+t} (\alpha_t - 1)^2 \quad (9)$$

where  $\lambda_j$  is the Lagrange multiplier of the multiple constraints.

We describe the process of SDFC algorithm to complex detection of time-dependent PPI network as Algorithm □.

## V. EXPERIMENTS

In this section, we apply our algorithm to a synthetic dataset and a real dataset. The synthetic dataset [22] contains 350 vertexes and 3 complexes, having 55, 110, 210 members, respectively. There are 20 temporal networks. For simplicity, we just show two temporal networks and the details of locality in this dataset in Fig. 2.

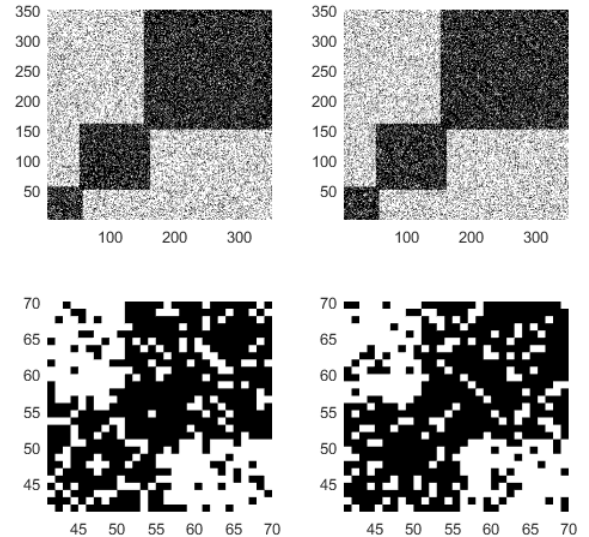


Fig. 2. The first and second temporal network of synthetic dataset and the local details of the two networks. We can see on the local details in these images that there is an explicit overlapping around the 55th node.

The real dataset is protein interactions of Yeast [23] [24]. We utilize dynamic Protein interaction networks construction method to construct 36 temporal networks, which has been proved effective to evolving process of protein by [25]. Details of the dataset is discussed in later of this section.

We compare our algorithms with four others. One algorithm is based on static PPI network, and we call it as StaticPPI [12]. Another two algorithms are based on dynamic network as being equal, and we call it as DynamicPPI\_1 [26]. Another algorithm of the two algorithms just consider the last temporal networks, and we call it Dynamic PPI\_2 [27]. The last algorithm is based on overlapping clustering, and the author named it as ClusterONE [7].

### A. Experiment on synthetic dataset

We carry on this experiment with our algorithm and other four algorithms. Since we know the real complexes of the synthetic data, we can use the Normalized Mutual Information (NMI) [22] to measure the accuracy of SDFC. NMI takes values in  $[0, 1]$ . NMI is equal to 1, if the complexes got from our algorithm are the same with real complexes. NMI is equal to 0, if complexes got from our algorithm are totally different from the real complexes.

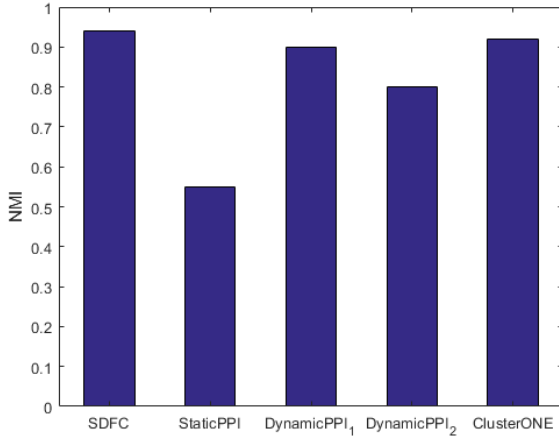


Fig. 3. NMI results of different algorithms on synthetic dataset.

As Fig.3 shows, the proposed SDFC algorithm deals with a PPI network as time-dependent network and finds overlapping groups. We choose the time window  $t = 5$  in this experiment. As shown, the results of our algorithm get the largest NMI value. The StaticPPI algorithm is the worst. Because it neglects too much information. It turns out that our algorithm considers time information and overlapping and gets accurate complex division for synthetic dataset.

### B. Experiment on real world datasets

For the real datasets, we don't have the ground truth and we take modularity [21] to measure the accuracy of our algorithm. This measure gives a value between  $-1$  and  $1$ . It was defined as  $Q = \sum_{ij}(A_{ij} - P_{ij})\delta(g_i, g_j)$  to describe the quality of the partition, where  $\delta(g_i, g_j) = 1$  if the complex assignments  $g_i$  and  $g_j$  of nodes  $i$  and  $j$  are the same and  $0$  otherwise and  $P_{ij}$  is the expected weight in our model. The higher the value of modularity, the better the result of complex detection is.

First, we need to construct the temporal networks according to [25]. A protein is considered to be expressed and appear at a time point only if its expression level at that time point is greater than the predefined threshold. Given a static PPI network [23] and the expression process [24], we can construct the temporal PPI networks. There are 2361 proteins in the PPI network. And 1358 proteins can be traced over time. Although we cannot get all the proteins' temporal information, we can also improve the quality of the results by using this information.

Then we want to find the value of time window we should take in this dataset. Because if we use a large time window, it

will introduce useless information in the dataset. However, if we choose a small value, we cannot handle the noise of the PPI network. We analysis the similarity between each time stamp. For the sake of simplicity, the similarity of last 10 time stamps is plotted as a heat map in Fig. 4(a). Each color section shows the inter-simplex relevance between two time stamps. For instance, the first column and the first row shows that the similarity of  $t_{27}$  itself is the most similar networks. In Fig. 4(b), we compute the similarity between the last time stamp and all the 36 time stamps. The result tells us that the nearest time stamp is the most similarity network and we choose 5 as the time window.

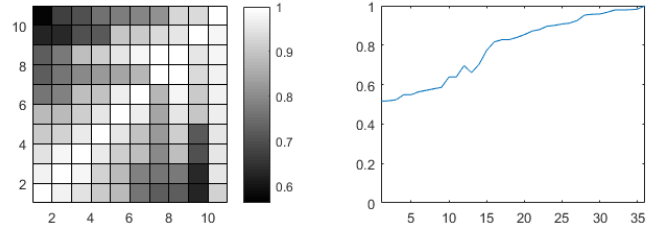


Fig. 4. (a) The heat map to illustrate the similarity between time-dependent networks. (b) The similarity of 36th time stamp with the other time stamps in the time-dependent network.

Another important problem is how to determine the number of clusters  $c$ . This parameter is the keystones of many clustering algorithms. We compute the modularity by setting  $c = 2$ . After that, when the algorithm get a local minimum, we increase  $c$  and optimized the cost function until it converges again. We keep on increasing the number of  $c$  until the value of modularity does not increase as the increasing of  $c$ . The influence of  $c$  is shown in Fig. 5.

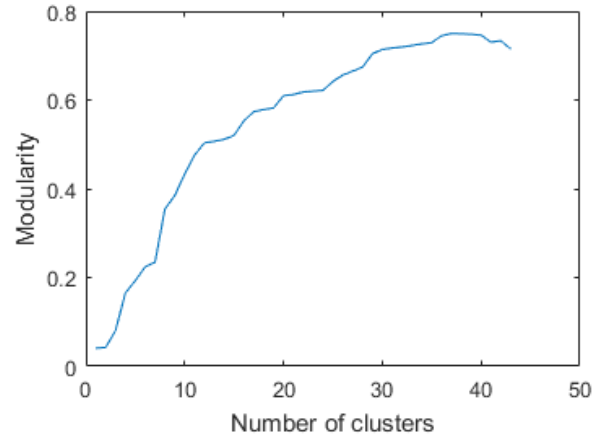


Fig. 5. The influence of number of clusters in detecting complexes.

Finally, we compare our algorithm with others. Fig. 6 reveals that our algorithm get larger modularity than other four algorithms. It turns out that our algorithm treating dynamic and overlapping PPI network are better than others in real world dataset.

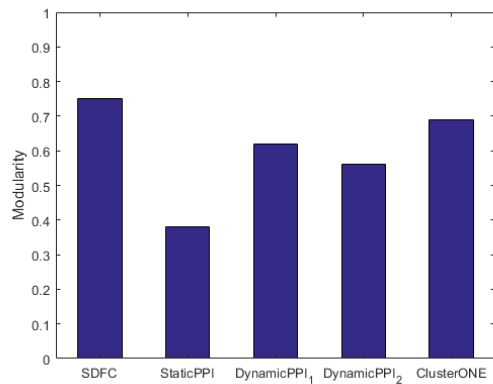


Fig. 6. Modularity results of different algorithms on real world dataset.

## VI. CONCLUSIONS AND FURTHER STUDY

In this paper, we propose a new model (IM-TDN) to describe time-dependent network and an algorithm (SDFC) to detect the overlapping protein complexes. We use the model to integrate time information to compute the similarity matrix. Then we analyze the influence of the parameters in our algorithm. We apply our algorithm on a synthetic data and a real dataset. The results show that our algorithm has higher accuracy in detecting complexes of PPI comparing with the competing algorithms.

In the future, we would like to extend this algorithm to utilize the multiplex information in the PPI networks. Because there are many different types of interactions between proteins. Different kinds of interactions play different function in PPI network. Consider more information can improve the accuracy of the division.

## ACKNOWLEDGMENT

This work was supported by National High Technology Research and Development Program of China (No.2013CB329605) (973 Program) and the Australian Research Council (ARC) under Discovery Grants DP140101366 and DP170101632.

## REFERENCES

- [1] A Vinayagam, J Zirin, C Roesel, Y Hu, B Yilmazel, AA Samsonova, RA Neumüller, SE Mohr, N Perrimon. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nature methods*. 2014 Jan 1;11(1):94-9.
- [2] M Li, W Chen, J Wang, FX Wu, Y Pan. Identifying dynamic protein complexes based on gene expression profiles and PPI networks. *BioMed research international*. 2014 May 18;2014.
- [3] B Chen, J Shi, S Zhang, FX Wu. Identifying protein complexes in protein-protein interaction networks by using clique seeds and graph entropy. *Proteomics*. 2013 Jan 1;13(2):269-77.
- [4] S Srihari, HW Leong. Parameterized Algorithms for Clustering PPI Networks. *arXiv preprint arXiv:1505.00925*. 2015 May 5.
- [5] EM Hanna, N Zaki. Detecting protein complexes in protein interaction networks using a ranking algorithm with a refined merging procedure. *BMC bioinformatics*. 2014 Jun 19;15(1):1.
- [6] O Maruyama, Y Kuwahara. RocSampler: Regularizing overlapping protein complexes in protein-protein interaction networks. In *Computational Advances in Bio and Medical Sciences (ICCBS)*, 2016 IEEE 6th International C.

- [7] T Nepusz, H Yu, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nature methods*. 2012 May 1;9(5):471-2.
- [8] B Cao, J Luo, C Liang, S Wang. Detecting overlapping protein complexes in weighted protein-protein interaction networks using pseudo-clique extension based on fuzzy relation. In *Neural Networks (IJCNN)*, 2016 International Joint Conference on 2016 Jul 24 (pp. 1244-1252). IEEE.
- [9] B Chen, W Fan, J Liu, FX Wu. Identifying protein complexes and functional modules—from static PPI networks to dynamic PPI networks. *Briefings in bioinformatics*. 2014 Mar 1;15(2):177-94.
- [10] IW Taylor, R Linding, D Warde-Farley, Y Liu, C Pesquita, D Faria, S Bull, T Pawson, Q Morris, JL Wrana. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology*. 2009 Feb 1;27(2):199-204.
- [11] GD Bader, CW Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003 Jan 13;4(1):2.
- [12] AJ Enright, S Dongen Van, CA Ouzounis. An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*. 2002 Apr 1;30(7):1575-84.
- [13] G Getz, E Levine, E Domany, MQ Zhang. Super-paramagnetic clustering of yeast gene expression profiles. *Physica A: Statistical Mechanics and its Applications*. 2000 May 1;279(1):457-64.
- [14] B Cao, J Luo, C Liang, S Wang. Detecting overlapping protein complexes in weighted protein-protein interaction networks using pseudo-clique extension based on fuzzy relation. In *Neural Networks (IJCNN)*, 2016 International Joint Conference on 2016 Jul 24 (pp. 1244-1252). IEEE.
- [15] J Wang, J Ren, M Li, FX Wu. Identification of hierarchical and overlapping functional modules in PPI networks. *IEEE transactions on nanobioscience*. 2012 Dec;11(4):386-93.
- [16] L Ou-Yang, DQ Dai, XL Li, M Wu, XF Zhang, P Yang. Detecting temporal protein complexes from dynamic protein-protein interaction networks. *BMC bioinformatics*. 2014 Oct 4;15(1):335.
- [17] IW Taylor, R Linding, D Warde-Farley, Y Liu, C Pesquita, D Faria, S Bull, T Pawson, Q Morris, JL Wrana. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology*. 2009 Feb 1;27(2):199-204.
- [18] G Chowdhury. Introduction to modern information retrieval. Facet publishing; 2010 Jul 31.
- [19] A Mollgaard, I Zettler, J Dammeyer, MH Jensen, S Lehmann, J Mathiesen. Measure of Node Similarity in Multilayer Networks. *PLoS one*. 2016 Jun 14;11(6):e0157436.
- [20] PN Pham, HT Nguyen, V Snael. Improving Node Similarity for Discovering Community Structure in Complex Networks. In *International Conference on Computational Social Networks 2016 Aug 2 (pp. 74-85)*. Springer International Publishing.
- [21] R Xu, D Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*. 2005 May;16(3):645-78.
- [22] L Tang, H Liu. Uncovering cross-dimension group structures in multi-dimensional networks. In *SDM workshop on Analysis of Dynamic Networks 2009 (pp. 568-575)*. ACM.
- [23] D Bu, Y Zhao, L Cai, H Xue, X Zhu, H Lu, J Zhang, S Sun, L Ling, N Zhang, G Li. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic acids research*. 2003 May 1;31(9):2443-50.
- [24] BP Tu, A Kudlicki, M Rowicka, SL McKnight. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science*. 2005 Nov 18;310(5751):1152-8.
- [25] J Wang, X Peng, W Peng, FX Wu. Dynamic protein interaction network construction and applications. *Proteomics*. 2014 Mar 1;14(4-5):338-52.
- [26] S Brohee, J Helden Van. Evaluation of clustering algorithms for protein-protein interaction networks. *BMC bioinformatics*. 2006 Nov 6;7(1):488.
- [27] S Asur, D Ucar, S Parthasarathy. An ensemble framework for clustering protein-protein interaction networks. *Bioinformatics*. 2007 Jul 1;23(13):i29-40.