Identification of Protein Complexes in PPI Networks by a Fuzzy Method

Danling Wang
School of Mathematics and Physics,
University of Science and Technology Beijing, Beijing, 100083, China
Email: wang_dan_ling@sina.com

Yanfei Wang*
School of Sciences,
China Agricultural University, Beijing, 100083, China
Email: yfmu@sina.com

***Abstract*—In this paper, we introduce a novel fuzzy method which is combined fuzzy relation, interaction probability values and hub structure to detect sub-communities in complex networks. We apply our method on yeast protein-protein interaction network to identify the protein complexes. Compared with traditional method, more protein complexes have been identified by this new fuzzy method. Meanwhile, we evaluate our method on two real social networks. The experimental results show that our method works well and give a reasonable understanding of these communities.**

***Index Terms*—Complex networks, fuzzy clustering, sub-communities, protein complexes, PPI networks.**

I. Introduction

A community structure or module, which consists of groups of vertices with dense connections, is an important feature in complex networks. It plays a significant role in analysis of complex networks. Different communities often correspond to different functional organizations. [1] For instance, in protein-protein interaction (PPI) networks, sub-communities are topological expression of protein complexes; in social networks, we can snoop the groupuscule according to the sub-structure of networks. Therefore, detecting sub-communities can enhance our understanding of the whole network and its functional modules. In the past several decades, various methods for sub-community detection were proposed. Especially in bioinformatics, some famous algorithms were applied to identify protein complexes from PPI networks. The simplest representation of PPI networks takes the form of a mathematical graph consisting of nodes and edges. Proteins are represented as nodes and edges represent physical interactions between proteins. Protein complexes are groups of proteins that interact with each, so they are usually dense sub-communities in PPI networks. Therefore, density-based clustering methods are widely applied for identifying protein complexes. [1-3]. The most popular density-based clustering method is the Clique Percolation Method (CPM) proposed by Palla et al. [1] for detection of overlapping protein complexes as k-clique percolation clusters. A k-clique is a complete full connected sub-network of size k. Based on CPM, a powerful tool named CFinder for identifying overlapping protein complexes has been developed by Adamcsek et al. [2].

In general, less protein complexes can be identified for larger values of *k*. The authors of CPM suggest using the values of *k* between 4 and 6 to analyze PPI networks. However, mining fully connected sub-networks is too restrictive in dealing with real biological networks. There are many other topological structures that may represent a protein complex in a PPI network, such as the star shape, the linear shape, and the hybrid shape which are shown in Fig. 1. Therefore, if we just identify the fully connected sub-networks, we will miss lots of protein complexes with the shape described above and the amount of identified protein complexes will decrease.

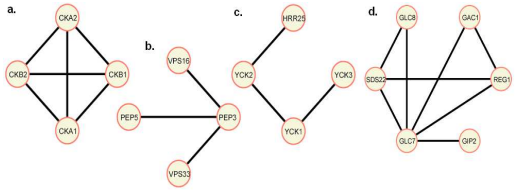

Figure 1. Projection of selected yeast MIPS complexes.

To overcome this problem, we combine the fuzzy relation clustering method with the graph model. Since the fuzzy set theory was proposed by Zadeh in 1965, it has been applied in many fields[4-6]. Fuzzy relation can effectively describe the uncertainty information between two objects, like the concepts “similar” and “different” [7]. Thus we establish a fuzzy relation model between every pair of nodes in the network and use the operations of fuzzy relation to obtain sub-networks. However, we cannot ignore the original structure of the network which contains important information for clustering analysis. That’s why we consider the sub-networks obtained from fuzzy relation model as the skeleton and calculate the interaction probability of each node to identify the

doi:10.4304/jsw.8.12.3132-3138

overlapping and non-overlapping sub-networks. In these sub-networks, some protein complexes exist.

We applied the method on yeast PPI network and compared with the clique percolation method. For the same data, more protein complexes have been identified. We also applied our method on two social networks. The results showed our method work well for detecting sub-networks and give reasonable understandings of these social communities.

II. THEORETICAL BACKGROUD

A. Topological Properties of PPI Networks

It is important to describe the topological and dynamic properties of various biological networks in a quantifiable manner. The literature on topological analysis of real networks is vast; therefore in this chapter we just give a briefly discussion on the related concepts and properties. Comprehensive reviews can be found in[8][9]. Here, we give an example of one part of the yeast PPI network in Fig. 2 by which we can understand these concepts better.

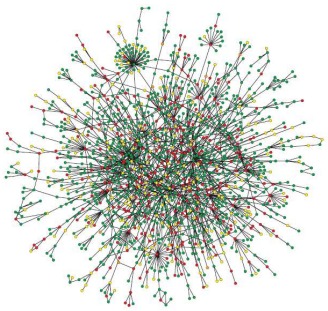

Figure 2. An example of protein-protein interactions network in yeast. from www. visualcomplexity .com

Definition 1 A graph (or network) is a ordered pair $G = (V, E)$, where
(i) $V = \{v_1, v_2,\ldots, v_n\}$, $V \neq \varnothing$, is called the vertex or node set of G;
(ii) $E = \{e_1, e_2,\ldots, e_m\}$ is the edge set of G in which $e_i = \{v_j, v_t\}$ or $<v_j, v_t>$ is the edge linking two nodes v_j and v_t.

Definition 2 The two nodes linked by one edge are called adjacent nodes; the edges linking these node are called adjacent edges.

Networks are naturally represented in matrix form. A graph of N nodes is described by an $N \times N$ adjacency matrix A whose non-zero elements a_{ij} indicate connections between nodes i and j. For undirected networks, a non-diagonal element a_{ij} of an adjacency matrix is equal to the number of edges between nodes i and j, and so the matrix is symmetric. In our method, adjacency matrix is used to calculate the similarity between two different nodes.

Definition 3 The edge clustering coefficient [10] is defined as the number of triangles which really include this edge divided by the number of all triangles which possibly include this edge. Let K_u and K_v be the degrees of nodes u and v respectively. Then the clustering coefficient of the edge linking u and v is

$$C_{u,v}^{(3)} = \frac{Z_{u,v}^{(3)}}{\min\{K_u - 1, K_v - 1\}},$$

where $Z_{u,v}^{(3)}$ means the number of triangles built on the edge. However, this definition is not feasible when the network has few triangles. Errors will occur when the number of possible triangles is zero. To avoid this limitation, Sun et al. [11] modified the definition of edge clustering coefficients by calculating the common neighbours instead of the triangles. Thus a new definition of edge clustering coefficient is given:

$$C_{u,v} = \frac{\left|N_v \bigcap N_u\right| + 1}{\sqrt{N_v \bullet N_u}}$$

where N_v and N_u represent the sets of neighbours of nodes v and u respectively. $C_{u,v}$ is a local variable; it quantifies how similar the two nodes v and u are connected by the edge $e_{u,v}$. If there is no edge between node v and u, then we consider $C_{u,v} = 0$. If v and u are the same node, then we let $C_{u,v} = 1$. From the definition we can see that the larger the value is, the more similar the two nodes are. In our method we use $C_{u,v}$ to calculate the similarity value between two proteins in PPI networks and transfer the adjacent matrix of PPI networks into a similarity matrix. We then use the fuzzy relation method in clustering analysis to find the sub-networks, which are possibly protein complexes in PPI networks.

B. Fuzzy Relation

Fuzzy relation is also proposed by Zadeh. In this section we will give some introduction on fuzzy relation theory. The letter 'R' can denote not only a fuzzy relation, but also a fuzzy matrix based on the fuzzy relation.

Definition 4 Let U and V be nonempty sets. A fuzzy relation $R \in F(U \times V)$ is a fuzzy set of the Cartesian product $U \times V$, $F(U \times V)$ is the set of all the fuzzy relations of $U \times V$.

Before doing clustering analysis based on fuzzy matrix, we have to make sure the fuzzy relation is a fuzzy equivalence relation. Here, we give the definition of fuzzy equivalence relation and fuzzy equivalence matrix.

Definition 5 Let $R \in F(U \times U)$. R is a fuzzy equivalence relation if it satisfies the following conditions:

(1)Reflexivity: $\forall u \in U, R(u,u) = 1$;
(2)Symmetry: $\forall (u_i, u_j) \in U \times U, R(u_i, u_j) = R(u_j, u_i)$;
(3)Transitivity: $R \supseteq R^2$.

If U is finite, then the fuzzy relation R on U can be expressed by fuzzy matrix, which is called a fuzzy equivalence matrix.

Theorem 1 Let R be a fuzzy similarity matrix, then there is a smallest nature number k ($k \le n$) such that $t(R) = R^k$. On the other hand, for any l greater than k, we always have $R^l = R^k$.

The above theorem suggests $t(R)$ is a fuzzy equivalence relation and the fuzzy matrix based on it is a fuzzy equivalence matrix. We can transfer a fuzzy similarity matrix to a fuzzy equivalence matrix by computing the transitive closure $t(R)$. For simplicity, we use the method of squares to compute $t(R)$:

$$R \to R^2 \to R^4 \to ... \to R^{2k} \to ...,$$

If $R^i \circ R^i = R^i$, then R^i is the transitive closure $t(R)$.

Definition 6 Let $R = (r_{ij})_{n\times m}$, $\forall \lambda \in [0,1]$, we have $R_\lambda = (r_{ij}(\lambda))_{n\times m}$, where

$$r_{ij}(\lambda) = \begin{cases} 1, & r_{ij} \ge \lambda \\ 0, & r_{ij} < \lambda \end{cases},$$

We call R_λ the λ cut matrix of R, and if

$$r_{ij}(\lambda) = \begin{cases} 1, & r_{ij} > \lambda \\ 0, & r_{ij} \le \lambda \end{cases},$$

then we call R_λ the strong λ cut matrix of R. If this section, we apply strong cut set to transfer a fuzzy matrix to a Boolean matrix for clustering sub-networks.

Now we know that, by using fuzzy matrix to perform clustering, the fuzzy matrix should be a fuzzy equivalence matrix. In practice, mostly fuzzy matrices established are fuzzy similarity matrices, thus we need to compute its transitive closure by the method of squares. After we obtain its transitive closure, we need to transfer it to a Boolean matrix by computing its λ -cut matrix. Here we give more details about how to use fuzzy relation matrix to perform clustering analysis. The steps of fuzzy clustering are as follows:

1. Data normalization.
2. Establishing fuzzy similarity matrix.

Remark 1: After normalization of observation values, we can establish fuzzy similarity matrix via computing the similarity relation between any two samples. For different node i and j, we compute the similarity value between them, which should satisfy $0 \le r_{ij} \le 1$, i, j = 1, 2, …, n. Then we obtain a fuzzy similarity matrix R which shows the similarity between every pair sample:

$$R = \begin{pmatrix} r_{11} & r_{12} & ... & r_{1n} \\ r_{21} & r_{22} & ... & r_{2n} \\ ... & ... & ... & ... \\ r_{n1} & r_{n2} & ... & r_{nn} \end{pmatrix}.$$

Remark 2: In our problem we apply the clustering coefficient defined by Sun et al. [11] based on the interaction matrix of a network which we have introduced in Definition 3.

3. Computing the transitive closure of the fuzzy similarity matrix via the method of squares.
4. Transforming the transitive closure to a Boolean matrix via computing the λ -cut matrix. The Boolean matrix is the skeleton of clustering result.

III. METHODS

In this part we introduce our method on identification of protein complexes. We combine fuzzy relation clustering analysis with IP value and hub structure in sub-networks, which we call the FRIPH method.

We can obtain the cluster skeleton of a PPI sub-network via the Boolean matrix transformed from the transitive closure of a fuzzy similarity matrix. Some protein complexes may be in these clusters. However, some protein complexes are overlapping on each other, which means each protein may be involved in multiple complexes. This is particularly true for protein interaction networks for most proteins having more than one biological function. For instance, there are 2750 proteins in the CYGD database [12], however the amount of protein complexes is 8931. Thus, it is very significant to identify overlapping protein complexes. Li et al. [13] proposed a new concept, Interaction Probability IP_{vi}, to measure how strongly an outside vertex v connects to another sub-network which doesn't contain v. Interaction probability IP_{vi} of any vertex v with respect to any sub-network i of size $|V_i|$ is defined as

$$IP_{vi} = \frac{|E_{vi}|}{|V_i|},$$

where $|E_{vi}|$ is the number of edges between the vertex v and the sub-network i. As shown in Fig. 3 below, the IP_{vi} of the vertex v to the sub-network i is 0.5.

For every vertex v in the original PPI network, we calculate its IP_{vi} in all sub-networks, i =1, 2, 3, …, m. Suppose vertex v is in sub-network j. If sub-network i has the greatest IP_{vi} with vertex v, then v can be

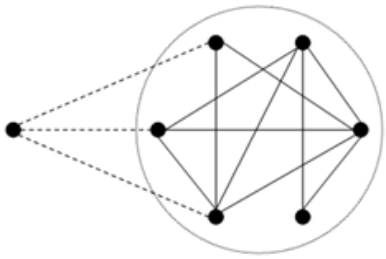

Figure 3. The interaction probability IPvi of a vertex v with respect to the sub-network i is 0.5

"added" to sub-network *i*, thus sub-network *i* will overlap with sub-network *j*. To summarize,

$$If\ IP_{vi} = \max_{k}(IP_{vk}), k = 1,2,...,m\ ,\ then\ v\ is\ also\ in\ sub\text{-}network\ i.$$

However, sometimes vertex *v* has the same greatest *IP* value with several sub-networks. In this situation, we need to compare the nodes connected to vertex *v* in these sub-networks. If vertex *v* is connected with a hub in sub-network *i*, then *v* can be also in sub-network *i*. Fig. 4 show a hub structure in PPI networks.

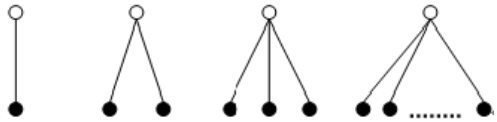

Figure 4. The hub structures in PPI networks.

The algorithm FRIPH can be divided into the following steps: 1. Generate an adjacency matrix from PPI data. 2. Choose a suitable method to compute similarity between each node in the network. 3. Compute transitive closure of the fuzzy matrix. 4. Transform the transitive closure to Boolean matrix via λ-cut matrix. 5. Compute IP values and compare hub structure in the original network to make sub-networks overlap.

IV Result and Discussion

Firstly, we apply our method to two social networks. The first one is Zachary's karate club network. The second one is network of American college football teams. We aim to identify the non-overlapping sub-networks in the two networks. We use visualized the networks structure by "Cytoscape" which is a network visualization software and can be download from the website www. cytoscape .org.

A. Zachary's Karate Club Network

This is a widely used data as a test example for methods of identifying sub-networks in complex networks. In this data, there are 34 nodes representing 34 people. Zacahry observed them for more than 2 years. During this study, a disagreement developed between the administrator (node 34) of the club and the club's instructor (node 1), which ultimately resulted in the instructor's leaving and starting a new club, taking about a half of original club members with him. Zachary constructed the network between these members in the original club based on their friendship with each other and using a variety of measures to estimate the strength of ties between individuals. Fig. 5 shows the graph of the network. There are 78 edges and two non-overlapping sub-networks in the graph, representing two groups of people with the administrator (circle label) and the instructor (square label). We apply our FRIPH to try to identify the two groups.

Following the step of FRIPH described in Figure 4.8, we separated the original networks into two sub-networks and two single nodes when we choose the value of λ as 0.75. Fig. 6 shows the result we obtain.

Comparing Fig. 6 with the original network in Fig. 5, the instructor group is perfectly separated from the original network. For the administrator group, node 10 and node 28 are not in the group but as two single points. The remaining nodes are all in administrator's group. Then we calculate the IP values of node 10 and node 28. For node 28 in Figure 5, it is connected with nodes 34, 24 and 25, which all belong to administrator's group; only node 3 belongs to instructor's group, thus the IP value of node 28 in administrator's group is greater than that of node 28 in instructor's group. Node 28 should belong to administrator's group. For node 10, it is just connected with nodes 34 and 3. However, node 34 is the administrator which is the hub of that group. Therefore, node 10 also belongs to administrator's group. From the result of karate club data, the FRIPH method detects the two sub-networks correctly. However, the edges in the sub-networks are totally changed; these new edges have no meaning in the sub-network. But they have no effect on the correctness of groups of sub-networks.

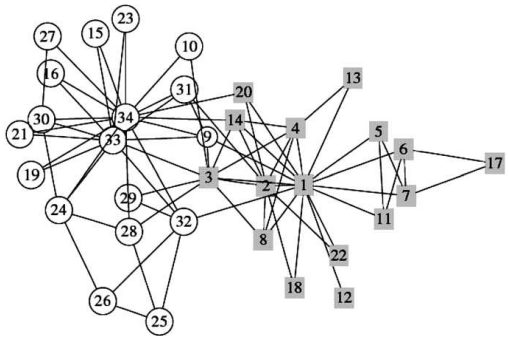

Figure 5. Zachary's karate club network.

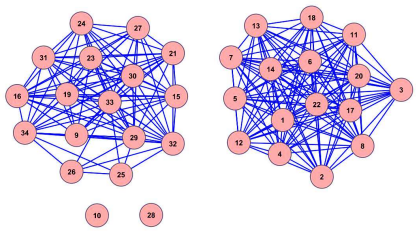

Figure 6. Sub-networks of Zachary's karate club network, obtained by FRIPH

B. American College football teams Network

The second social network we test is the network of American college football teams which represents the game schedule of the 2000 season of Division I of the US college football league. In this data set, there are 115 nodes representing the teams and 613 edges presenting games played in the course of the year. The teams are divided into 12 conferences containing around 8-12 teams each. We apply our method on this data set and obtain the result showed in Fig. 7. However, the result is not satisfactory. We make a comparison with the result of [14] which is considered as a good one and shown in Fig. 8. For our result most nodes in the last three sub-networks belong to the Sunbelt conference and should be in the same group of the grey points in Fig. 8, but they divide into three sub-networks and group with members of the Western Athletic conference. This happens because the Sunbelt teams played nearly as many games against Western Athletic teams as they did against teams in their own conference. Thus our method fails in this case. Meanwhile, there are 8 points which cannot be grouped in any sub-networks. In Fig. 8, the same problem exists and these points are shown in red colour. That's because these nodes generally connect evenly with more than one community, thus our method cannot group them into one specific sub-network correctly. These nodes are the "fuzzy" nodes which cannot be classified correctly by the current edge information. Generally, these points play a "bridge" role in two or more sub-networks of the original network.

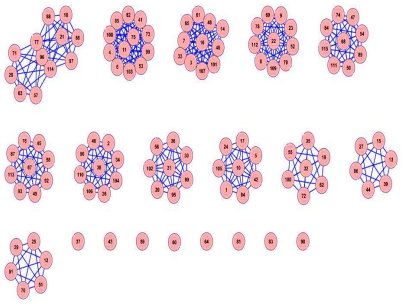

Figure 7. Sub-networks of American college football team network, obtained by FRIPH.

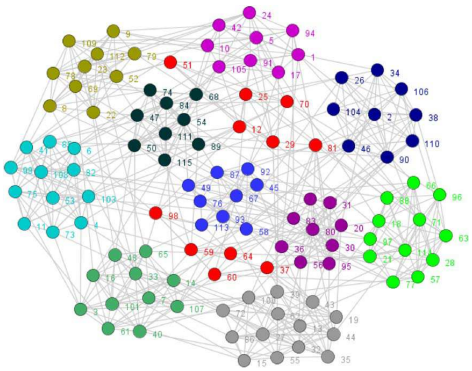

Figure 8. Sub-networks of American college football team network in [14].

C. Application to Identify Protein Complexes

Identification of protein complexes from PPI network is crucial to understanding principles of cellular organisation and predicting protein functions. Han et al. [9] have shown that sub-networks such as cliques and near-cliques indeed represent functional modules or protein complexes. Thus identification of sub-networks from a complex network becomes an important issue.

In this section, we apply our method on the protein interaction network of Saccharomyces cerevisiae, which was downloaded from the MIPS database [15] and make a comparison with the popular software CFinderAfter removing all the self-connecting interactions and repeated interactions, the final network includes 4546 yeast proteins and 12319 interactions. The network diameter is 13 and the average shortest path length is 4.42. According to theannotate in MIPS database for Sacchromyces cerevisiae, there are 216 protein complexes identified by experiment, which consist of two or more proteins. The largest complex contains 81 proteins, the smallest complex just contains 2 proteins and the average size of all the complexes is 6.31.

To evaluate the effectiveness of FRIPH for identifying protein complexes, we compare the predicted clusters with known protein complexes in the MIPS database. There are 216 manually annotated complexes which consist of two or more proteins. We use the scoring scheme which is also applied in [16]and [17] to determine how effectively a Predicted Cluster (*Pc*) matches a Known Complex (*Kc*). The overlapping score between a predicted cluster and a known complex is calculated by the following formula:

$$OS(Pc, Kc) = \frac{i^2}{|V_{Pc}| \times |V_{Kc}|},$$

where i is the number of nodes which are the intersection set of size of predicted cluster and known complex, $|V_{Pc}|$ is the size of predicted sub-network and $|V_{Kc}|$ is the size of known complex. If a known complex does not have the same protein in a predicted sub-network, then the overlapping score is 0, and if they perfectly match with each other, the overlapping score is 1. A known complex and a predicted cluster are considered as a match if their overlapping score is larger than a specific threshold. The number of matched known complexes with respect to different overlapping score threshold is shown in Fig. 9 and Table 1

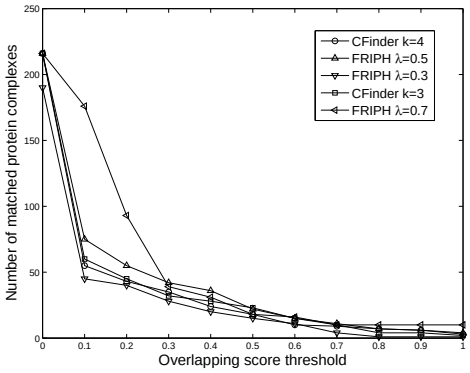

Figure 9. The number of known complexes matched by predicted sub-networks of FRIPH and CFinder with respect to different parameters and overlapping score.

As shown in Fig. 9 and Table 1, CFinder obtains best matching when k = 3. The number of known complexes matched to the predicted sub-networks detected by CFinder using k = 3, 4, 5, 6 are 55, 43, 20 and 11 with respect to $OS(Pc, Kc)$ = 0.2. The number of matched protein complexes decreases as k increases. In the work of Zhang et al.[18] and Jonsson et al.[19], this result was also deduced. That's because when k is determined, CPM just identifies the complexes which contain k or more proteins. For the FRIPH method, when λ=0, the PPI network doesn't change and all the nodes are in the same group. As λ increases, the number of matched complexes increases. When λ = 0.9, FRIPH obtains the best result and the number of matched complexes whose overlapping score is larger than 0.5 is stable. That is because when λ is increasing, the number of single proteins is increasing, thus the protein complexes with 2 or 3 proteins can be found much easier out of the original PPI network.

TABLE I.
TYPE SIZES FOR CAMERA-READY PAPERS

Overlapping Score	CFinder (K=4)	CFinder (K=4)	FEIPH λ=0	FEIPH λ=0	FEIPH λ=0	FEIPH λ=0	FEIPH λ=0
0	216	216	1	190	216	216	216
0.1	55	75	1	45	60	176	216
0.2	43	55	0	40	45	93	104
0.3	35	42	0	28	32	39	40
0.4	24	36	0	20	28	31	28
0.5	18	22	0	15	23	18	20
0.6	10	15	0	11	15	16	20
0.7	9	11	0	4	10	10	20
0.8	7	7	0	1	4	10	20
0.9	6	6	0	1	4	10	20
1	3	4	0	1	2	10	20

V. CONCLUSION

In this paper, we proposed a novel method which combines the fuzzy clustering，interaction probability as well as hub structure to identify the overlapping and non-overlapping community structures in PPI networks, then to detect protein complexes in these sub-networks. Our method is based on both the fuzzy relation model and the graph model. Fuzzy theory is suitable to describe the uncertainty information between two objects, such as 'similarity' and 'differences'. Additionally, the original graph model contains significant clustering information, thus we can not ignore the original structure of the network, but combine it with the fuzzy relation model. We applied the method on yeast PPI network and compared with CFinder. For the same data set, eventhough the precision of matched protein complexes is lower than CFinder, we detected more protein complexes. We also tested our method on two social networks. The results showed that our method works well for detecting sub-networks and gives a reasonable understanding of these communities.

ACKNOWLEDGMENT

This work is supported by Chinese Universities Scientific Fund No. 2013XJ010 and the Fundamental Research Funds for the Central Universities No. FRF-TP-13-020A.

REFERECNES

[1] G. Palla, I. Derenvi, I. Farkas and T.Vicsek, "Uncoveing the overlapping community structure of complex networks in nature and society," Nature, vol. 435, pp. 814-818, 2005.
[2] B. Adamcsek, G. Palla, I. Farkas, I. Derenyi and T. Vicse, "Cfinder: locating cliques and overlapping modules in biological networks", Bioinformatics, vol. 22, pp1021-1023, 2006.
[3] E.Zotenko., K.S. Guimaraes, R. Jothi and T.M. Przytycka, "Decomposition of overlapping protein complexes: a graph theoretical method for analyzing static and dynamic protein associations", Algorithms for Molecular Biology, vol. 1, pp7-14,2006.
[4] Y.Wei and X. Wu, "A new fuzzy svm based on the posterior probability weighting membership", Journal of computers, vol. 7, No. 6, pp1385-1392, 2012.
[5] Q. Zhang and M. Zhong, "Using multi-level fuzzy comprehensive evaluation to assess reservoir induced seismic risk", Journal of Computers, vol. 6, No.8 pp1670-1676, 2011.
[6] S. Xu, C. Li, S. Jiang, and X. Liu, "Similarity measures for content-based image retrieval based on intuitionistic fuzzy set theory", Journal of Computers vol. 7, No.7, pp1733-1742, 2012.
[7] L.A. Zadeh, "Fuzzy sets", Information and Control, vol. 8, pp. 338-353, 1965.
[8] D. Chakrabatti, "Tools for large graph mining" PhD thesis, School of Computer Science, Carnegie Mellon University,2005.
[9] J.D. Han, D. Dupuy, N. Bertin, M.E. Cusick and M. Vidal, "Effect of sampling on topology predictions of protein-protein interaction networks", Nat Biotechnol, vol. 23: pp839-944, 2005.

[10] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Defining and Identifying Communities in Networks, Proc, Natl. Acad. Sci, USA, vol. 101, pp2658-2663, 2004
[11] P.G. Sun, L. Gao and S.S. Han, "Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks", Information Sciences, vol.181, pp1060-1071, 2011
[12] U. Guldener, M. Munsterkotter, G. Kastenmuller, N. Strack, J. Van Helden, C. Lemer, J. Richelles, S. Wodak, J. Garcia-Martinez, J. Perez-Ortin, H. Michael, A. Kaps, E. Talla, G. Dujon, B. Ander, J. Souciet, J. De Montigny, E. Bon, C. Gaillarding and H. Mewes, "CYGD: the comprehensive yeast genome database", Nucleic Acides Tes., vol. 33, pp364-368, 2005
[13] M. Li, J.X. Wang, C. Zhao and G. Chen, "Identifying the overlapping complexes in protein interaction networks", Int. J. Data Mining and Bioinformatics, vol. 4, pp112-137 2010.
[14] S.H. Zhang, X.M. Ning and X.S. Zhang, "Identification of overlapping community structure in complex networks using fuzzy c-means clustering", Physica Q, vol. 374, pp.483-490,2007
[15] H. W. Mewes, "MIPS: analysis and annotation of proteins from whole genome ", Nucleic Acids Res, vol. 34, pp. 169-172, 2006
[16] A.D. King N. Przulj and I. Jurisica, "Protein complex prediction via cost-based clustering", Bioinformatics, vol. 20, pp839-944, 2004.
[17] M. Altaf-Ul-Amin, K. 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, pp1-13, 2006.
[18] S.H. Zhang, R.S. Wang and X.S. Zhang, "Identification of fuctional moduls in a PPI network by clique percolation clusteiring", Comutational Biology and Chemistry, vol. 30, pp 445-451, 2007.
[19] P. Jonsson, T. Cavanna, D. Zicha and P. Bates, "Cluster analysis of networks geerated through homology:tomatic identification of important protein communities involved in cancer metastatis", BMC Bioinformatics, vol. 7, pp. 1-13, 2006

Danling Wang received her M.Sc degree from University of Science and Technology Beijing, in 2008 and Ph.D in applied mathematics from Queensland University of Technology, Australia in 2011. She is a lecture in University of Science and Technology Beijing now. Her main research interests include fractal and multifractal analysis on complex networks, chaos and fractals, bioinformatics.

Yanfei Wang received his M.Sc degree from Dalian Maritime University, in 2008 and Ph.D in applied mathematics from Queensland University of Technology, Australia in 2011.He is a lecture in China Agriculture University now. His research interests include fuzzy clustering, fuzzy evaluation, bioinformatics.