ISSN 2229-5518

Community Detection based on Relationships Clustering in Complex Network

Owais Muhammad, Phouvyrath PHENGSAVATH, Sajid Saleem, Hashmi Samar

ABSTRACT

Research on complex networks has always been a research hotspot in many fields. In particular, the discovery and clustering of community structures is of great significance to our understanding of network characteristics. With the continuous development and popularization of information technology, the world has propel into the network era, there are various reasons for complex systems in the real world can be referred by complex networks, such as transportation networks, collaborative networks, social networks, and biological networks. The nodes in the network represent objects, and the edges between nodes represent the degree of attribute association between objects. Based on this, the main work of this article is as follows.

Firstly, this paper analyzes and summarizes representative methods for the discovery of existing complex network communities, and conducts in-depth study of link clustering, spectrum bi-section method, cohesion method, and splitting method. We also use L.C as the main research object, and study the calculation methods of node attribute expression, partition density, Dendron structure and other parameters, and realize the initial construction of the network community structure by dynamically adding and removing network edges. Secondly, aiming at the shortcomings of existing L.C clustering methods for overlapping community discovery, an Extended Link Clustering method (E.L.C) is proposed. The method uses the transformation matrix to express link similarity, and judge the incremental changes in the results of different community divisions based on a comparison of Dendron structure. Finally, in order to overcome the fuzzy effect of partition density on the hierarchical clustering, we propose a new modularity calculation method. It uses the density deviation statistics of the divided sub-networks to dynamically determine the optimal partition termination conditions and improve the efficiency and accuracy of E.L.C partitioning.

The experimental datasets includes karate data sets, dolphin data sets, U.S. political book data sets, U.S. college football league data sets, and protein network data sets. The simulation results show that compared with L.C and C.P.M community division methods, the E.L.C method proposed in this paper can quickly identify the community structure in different data sets. In addition, contrasting the termination conditions of community and the quality of overlapping community discovery, the E.L.C method is superior to the contrast method in accuracy and quantity.

KEY WORDS: Link Clustering, Community Detection, Overlapping Community Detection

Introduction:

With the rapid of evolution and progression communication network technology and high performance computing, it has more widespread coverage and pays more consideration to people's daily lives and social performance. Similarly, taking into an account of the transitive relation in a social connection between them, community like the people living in the same area also people share common characteristics or interests and network can determine a significant amount of potentially useful information. In the social network, a node represents an individual and edge between a pair of nodes represents the presence of a relationship, it will bring some data transfer skill by using network link nodes. Social network analysis over and over again aims to detect a variety of clustering of nodes in a network, nodes of high connectivity the means by which individual terminals and closely connected subgroups as well as subjectively there are distribute the roles to node based on the result of detection. Evaluating the characters of a node in community is helpful to supportive the social organization. Also defining feedback, aggregation feature to reproduce polymorphic and transmission process, computation of social associations cognitive progression of basically by extracting the key factors, in social explores the relationship there have tendency communication behaviours for people the same interests relationship to each other [1]. However, the role of a node is assigned in accordance to its rank without considering the community structure. The theory of graph be capable of traced back in 1736, Euler productively solved the problem of seven bridges in Konigsberg[2], with the more study of the graph other people can comprehend the character of the graph, because the network use the graph model to carry out do research. The network analysis began in 1930s and grow to be main do research topic of sociology [3]. Also in a social network, biological network, information network and so on, graph theory has the widespread application.

In the background of network is called the community structure[4], there has been an raised attention in solving the community structure, for example so-called "small world" means association of complex network has the characteristics of short path length and large clustering coefficient have differential factor, and the average path length is small as shown in figure 1-1. The network in figure contains communities these are corresponds to graph, because of key node were all linked each other nodes and just fill step, this is the classic concept idea degrees of sub-population. It plays an important role in supportive the behaviour of the system and the attributes of individuals, for people to understand network topology and function structure to help provide support for the use and revolution of network. The community network structure also called cluster or module, the communities densely linked nodes with sparse links relating between them, in different cases algorithm determine by community. By the way, how to know the community structure existing in the network has been widely apprehensive by exerts and scholars. It's the result of algorithm but it doesn't have perfect explanation by means. In traditional, the communities are seen as nodes same nature or playing similar roles in the diagram.

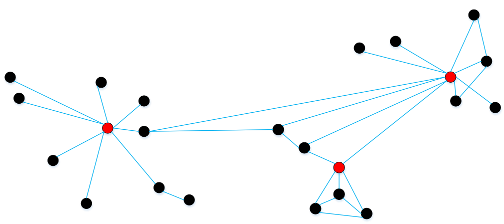


Figure 1 Small-world networks

In biology protein-interactive networks be capable of viewed as nodes and communities are made up of groups of proteins with special functions in the cells[5], in the Internet, a community be capable of viewed as a form of web pages dealing with the same or related topics[6]. In the community nodes can be divided into two types based on the location information of the nodes, one is a node in the central position, this type of node will be connected with the nodes of other communities, and also play an important role in control and stability. The second category, a type of node that acts as a bridge between different communities, which one its complexity in partitioning is likely to happen at the same time as a different community. This reflects the relationship between the communities and has a general understanding of the original diagram. Such classifications are of practical significance in social networks, and the process structure of a network system is of basic importance decisions to the performance of computing network[7], here we can discuss the hierarchical relationships of many network systems in the real world, such hierarchies are made up of communities, large communities comprise of small communities, and small communities encompass smaller communities, which form a hierarchical relationship through such happen again.

Research status

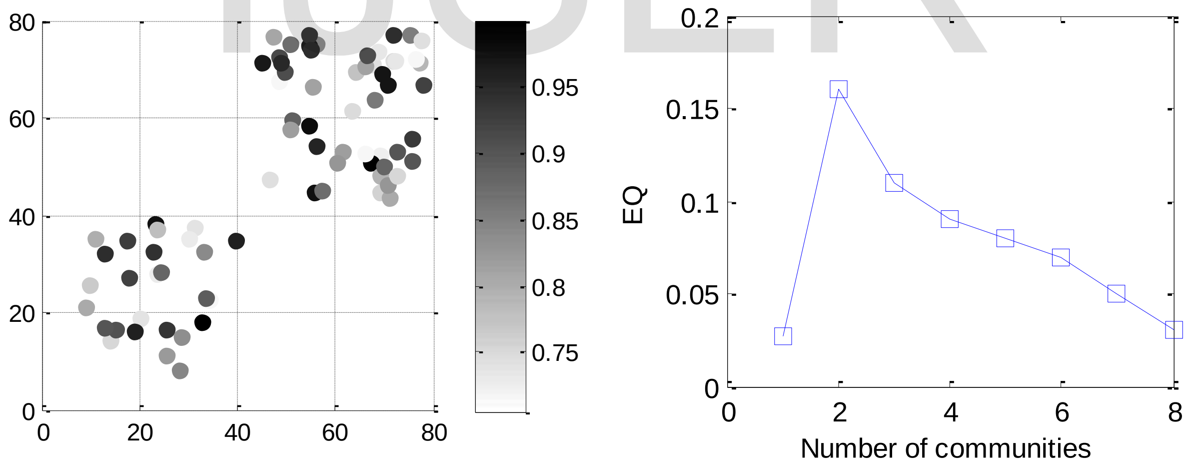
The aim of community detection is to find the modules in the network and if achievable to discover the hierarchical clustering relationship of the community structure according to the topology of the network, in fact the probability of a connection between nodes in the same community is greater than the nodes between different community, the functional networks communities as a topological attribute of a network it required to discover a community structures for analyze the relationship between the community structure and the overall structure of the network, the analysis of community structure first began with Weiss and Jacobson. After that new region of research and revealing advantage from a space in which a community can regard key question in network [8], they make the membership matrix of the organization and look for a working group by removing members that have relationships with dissimilar working groups, the idea of a link between the working group has set up the groundwork for some progressive community detection methods. The overlap in the community firstly discovered by Palla proposed a factional filtering algorithm clique percolation method (C.P.M). This paper proposes a classification of connection model based on clique percolation method the original method C.P.M[9], C.P.M method for community detection on overlapping, later Zhang proposed communities detection method called the C.-means[10]. The concept competition by random overlapping communities' detection in complex networks with clustering networks based on the idea of population competition. The connect similarity measures able assistance the greedy optimization process vertices into groups we able discover the local structure relationship[11]. Zhang proposed a new regularized spares random graph model for protein interaction networks to discover overlapping and various structural functional units representing interactions between multiple cells[12]. The agglomerative hierarchical algorithms such as Newman algorithm, mainly consider the hierarchical relationship of clustering dendrogram. The varying method used to overlapping community detect which are for the most part targeted at nodes such as the C.P.M method mentioned above, in general accepted focus of a community defined as a node of functional or similar nature, according to this definition, the community partitioning method does not consider the hierarchy and overlapping structure between these nodes. The overlap community detection based on analysis of line graph [13]. Ahn choose the Jaccard distance between the nodes as a measure of the two connections to connect the cluster for analyzing links, this method is called the link clustering (L.C) [14], they productively solved the problem in 2010 and delivered a new characteristic reference for overlapping community studies. Kalinka package linkcomm called R.-language basic on the link cluster for social network clustering[15], the software package withdraws a very helpful tool for analyzing networks by joining clustering. The community detection study also helps the development of metrics used to community structures evaluate. Newman proposed the modularity Q function for a community structure, the maximum of Q value refers a good communities used to evaluate of community detection, but the communities has no more within links.Q is 0, it not able be used directly to the evaluation of complex networks. So K.app and co-worker put forward an in-group proportion (I.G.P) index based on predictive accuracy for the calculation of cluster quality[16]. Shen and others put further E.Q called an extended modular degree address limitations of. Q[17], and A.hn proposed evaluating partition densities to detection complex networks clustering by link[18].

Experiment Network:

The advancement of extended link clustering its presentation can be compare with original method, on the real world network we choose five networks dataset, on these dataset are the most important fields on community detection including network karate club dataset, network Dolphin dataset, network American book politics dataset, network football dataset, interaction protein Y.2.H dataset by introduced under;

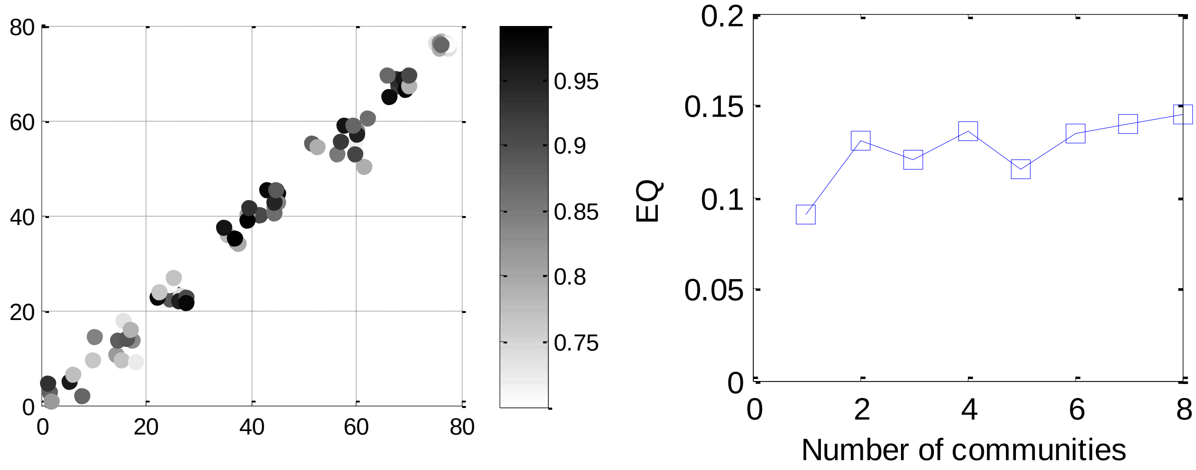
1）Karate club datasets

At an American college have a karate club namely Zhachary' karate club is shown the relationship network linking 34 friends, it is dataset interacting the most simple in the field social network analysis. The difference of opinion two leader between club president and instructor that is divide club into two groups. However, the members of clubs don't meet at class, but their keep relationship outside the club, their network dataset has 2 groups, 34 members (the members refer to nodes) and 78 relating (the relationship represent to links).



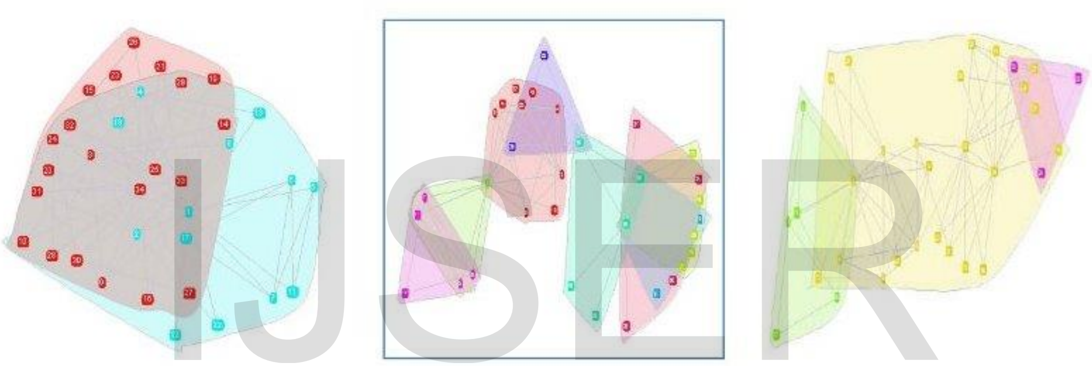
(a) (b)

Figure 1) E L C similarity matrix and graph



(a) (b)

Figure. 2) L C similarity matrix and graph



(a) E L C (b) L C (c) C P M

Figure 3 E L C, L C, C P M results in the diagram of Karate club datasets

Table 1) E L C, L C, C P M results in the Karate Club dataset

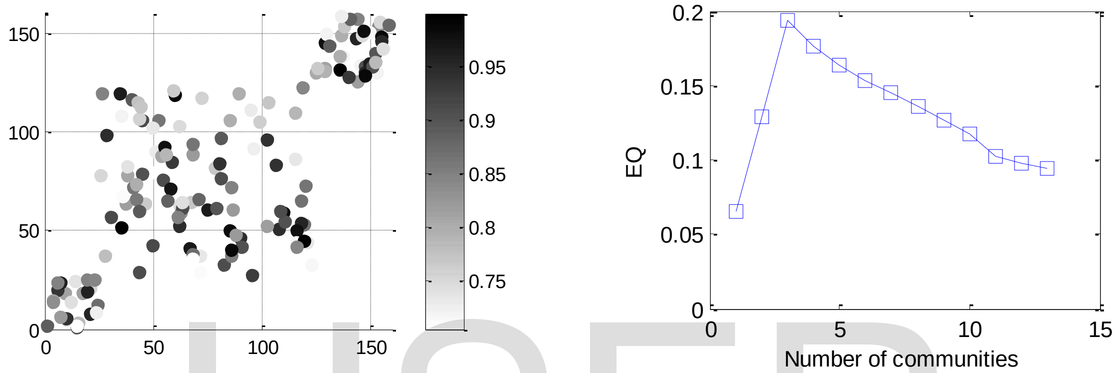
	I G P	P D	E Q
L C	0. 275	0. 285*	0. 145
C P M	0. 347	0. 201	0. 115
E L C	1*	0. 007	0. 160*

Note: the data shown good value in similar place marked with (*).

Zachary’s Karate Club dataset, the standard is divided into two categories, as described above have the core of two communities. From the similarity matrix as shown in figure 1 (a) and figure 2 (a), we able see that, the L C method denser less than the one produced by the transform matrix calculated by E L C method. As shown in figure 1 (a) these are two obvious groups and the transform matrix blocks of L C has less obvious. As shown in figure 3 (a) and figure 3(b), an E Q value is 0. 160 with 2 groups identified by E L C method, when an E Q value is 0. 145 With 8 groups identified by L C method, for L C method didn’t obtain the real world and only produced smaller groups. As shown in figure 3 (b) the value of E Q got 0. 115 with 3 groups divided by the C P M method and lower than E Q value of E L C method. When the biggest block in the network has found by the C P M method the final results is left two nodes uncovered.

2）Dolphin datasets

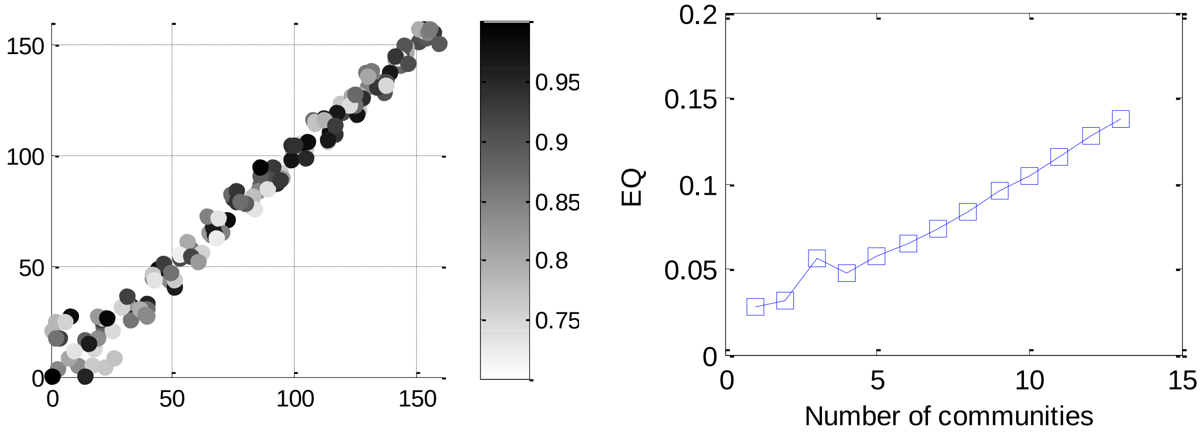
The communities dataset of dolphin by over seven years has been study the immigration and emigration of bottlenose dolphin's interaction each dolphin live in large location, that unclear sub relationship within sex act of dolphin, which studied between within sexes associations are the main factor of community analysis, their network dataset has 2 groups indentify by gender with 62 nodes (the nodes refer to the dolphins) and 159 links (the links is interactive dolphins during each of the two linked).



(a) (b)

IJSER

Figure 1) E.L.C similarity matrix and graph



(a) (b)

Figure 2) L.C similarity matrix and graph

(a) E.L.C (b) L.C (c) C.P.M

Figure 3) E.L.C, L.C, C.P.M results in the diagram of Dolphin datasets

Table 1) E.L.C, L.C, C.P.M results in the Dolphins dataset

	I.G.P	P.D	E.Q
L.C	0. .030	0. .030	0. .138
C.P.M	0. .063	0. .265	0. .182
E.L.C	0. .700	0. .092	0. .194

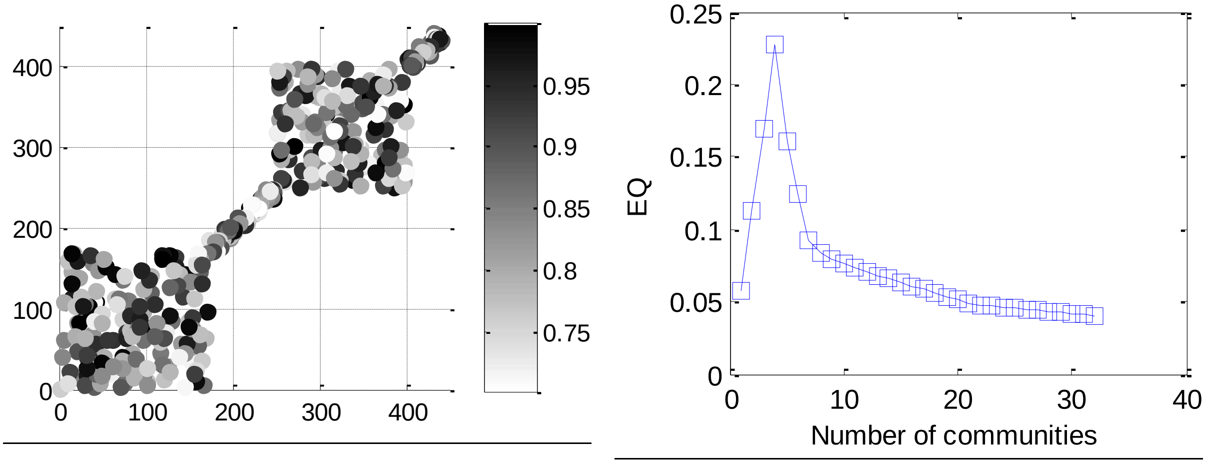
IJSER

As shown in figure 1 (a) and figure 2 (a) the L.C is unclear the transform matrix, but the transform matrix of the E.L.C clearly and produce 3 big clusters divided into network the E.Q value get 0. .194, and the E.Q value is 0.138 with 13 groups identified by the L.C method. After that only obtain 8 nodes with the biggest group, the E.Q value is 0. .182 with 4 groups identified by the C.P.M method.

3）Books about U S politics community

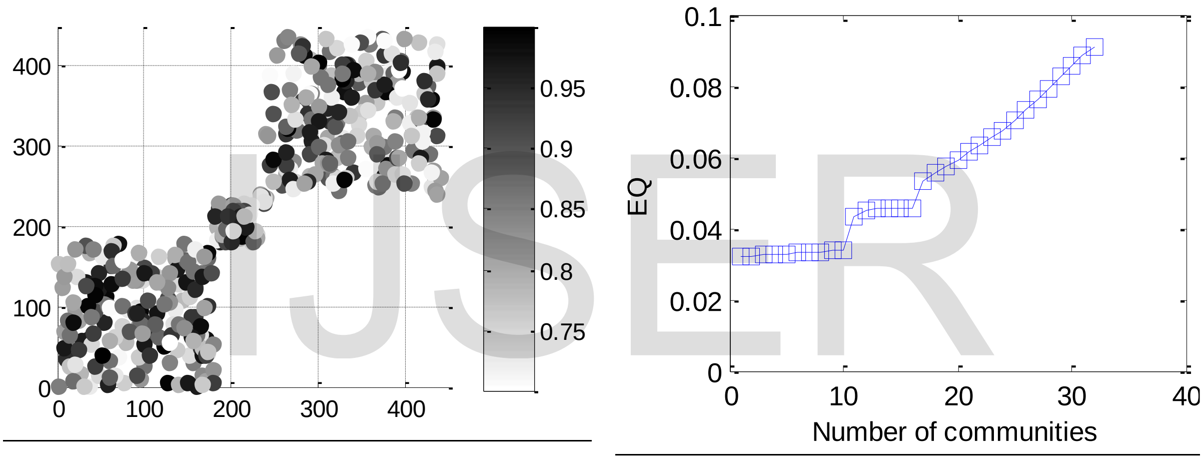
The network books dataset of American politics is a collect by Krebs and sold books on website Amazon.com booksellers. Krebs split kind of books to obvious conservative and liberal grouping. However, some title books have no cluster, for that reason Newman consider make three grouping of books namely conservative, liberal and neutral. In these network dataset the books bought on-line refer to the nodes and the link represent of two books are buy the same customer. Their network dataset has 3 groups among 105 nodes through 441 links.

ISSN 2229-5518



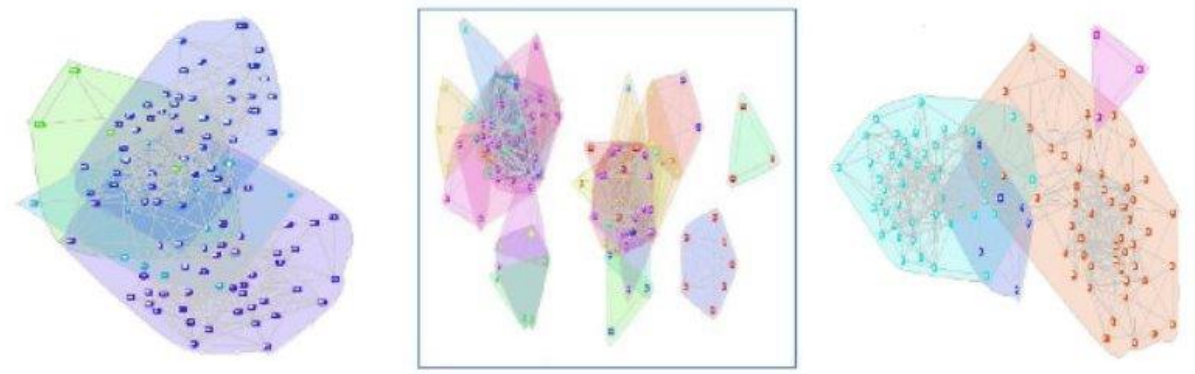
(a) (b)

Figure 1) E L C similarity matrix and graph



(a) (b)

Figure 2 L C similarity matrix and graph



(a) E L C (b) L C (c) C P M

Figure 3) E L C, L C, C P M results in the diagram of Books about U S politics

Table 1) E L C, L C, C P M Books about U S politics community datasets

	I G P	P D	E Q
L C	0. 078	0. 287*	0. 091
C P M	0. 278	0. 148	0. 221
E L C	0. 563*	0. 136	0. 227*

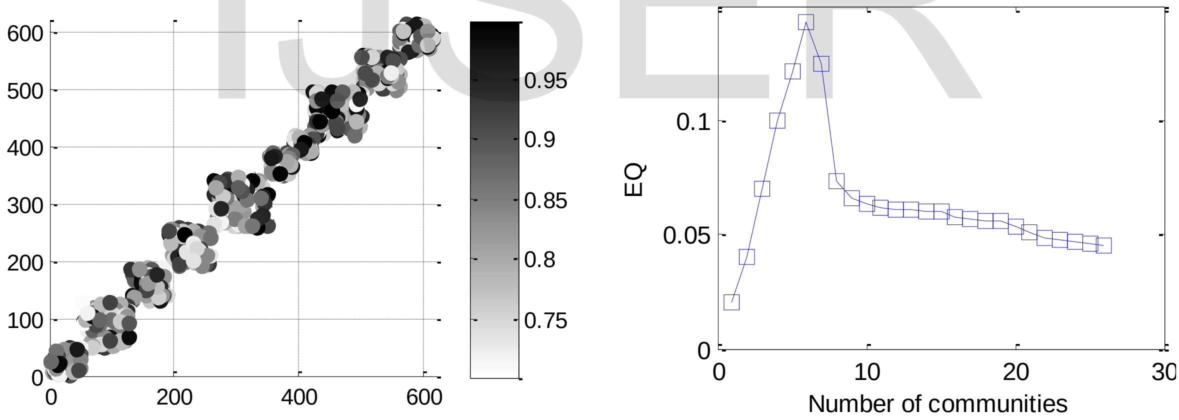
Note: the data shown good value in similar place marked with (*).

As shown in figure 1 (a), and figure 2 (c), the blocks are relatively and relatively sparse unclear by the L C method is transform matrix. Other while, the transform matrix by E L C generates that clearly and shows 2 big groups. The E Q value is 0. 228 with 4 groups identified by the E L C method. In the other hand, the E Q value is 0. 091 with 32 groups identified by the L C method, and the C P M method identifier an E Q value is 0. 221 via 4 groups. One more time, the number of groups performs well in identifying by the C P M method.

4）N F L Datasets (American College Football) University

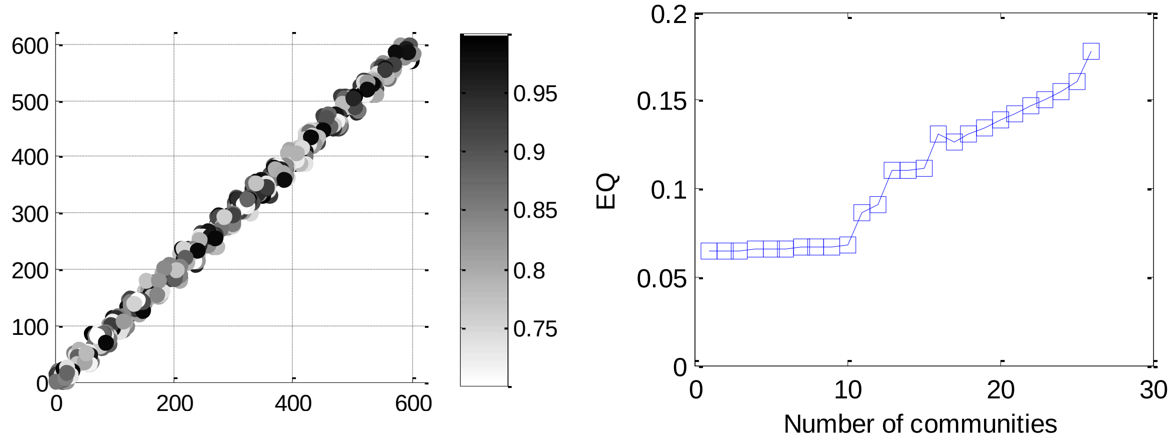
The timetable dataset of the football games had between 2000 of American school, which discussion the all team split out to 12 groups including during 8 to 12 of each team, in their network the each team refer to the nodes and among two team denote links. Regularly, the team at similar gathering have more links than team different meeting; their network dataset has 12 groups among 115 nodes and relation by 615 links.

IJSER



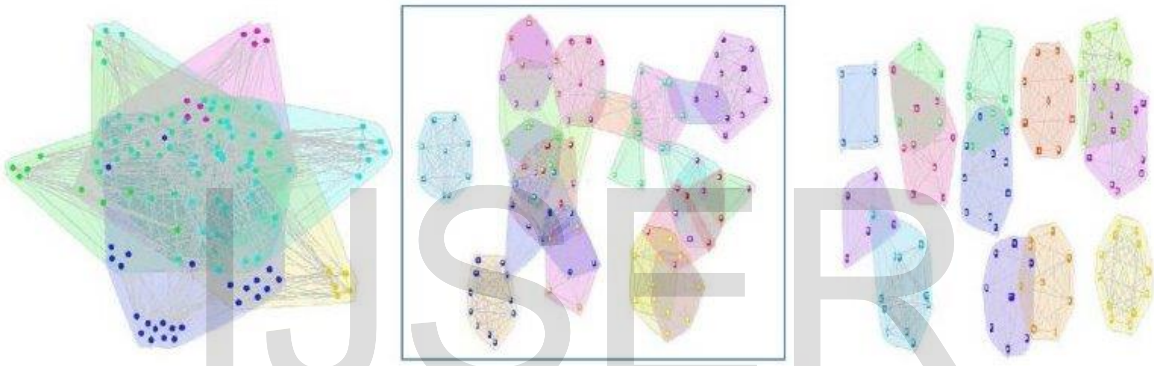
(a) (b)

Figure 1) E.L.C similarity matrix and graph



(a) (b)

Figure 2) L.C similarity matrix and graph



IJSER

(a) E.L.C (b) L.C (c) C.P.M

Figure 4-12 E.L.C, L.C, C.P.M results in the diagram of N.F.L datasets

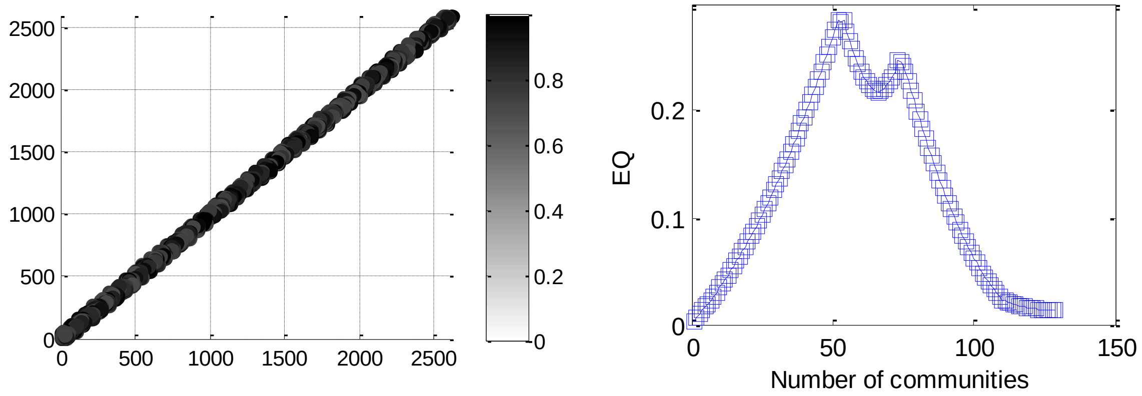
Table 1) E.L.C, L.C, C.P.M in the N.F.L Datasets

	I G P	P D	E Q
L.C	0. .173	0. .551*	0. .178
C.P.M	0. .036	0. .539	0. .283*
E.L.C	0. .500	0. .182	0. .143

Note: the data shown good value in similar place marked with (*).

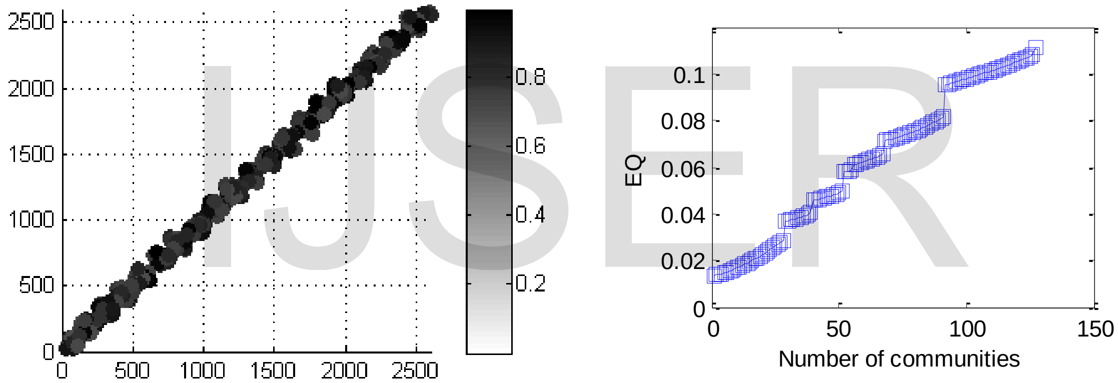
As shown in figure 1 (a) and figure 2 (a), the L.C method transform matrix distinguish the blocks unclear, but an E.Q value is 0. .143 with 6 groups achieved by the E.L.C method able distinguish almost 10 blocks. The EQ value is 0. .178 with 26 groups identified by the L.C method. the E.Q value is 0. .283 with 13 groups identified by the C.P.M method, as shown the LC method and E.L.C method lower than the one obtain by the C.P.M method.

5) Y.2.H Datasets



(a) (b)

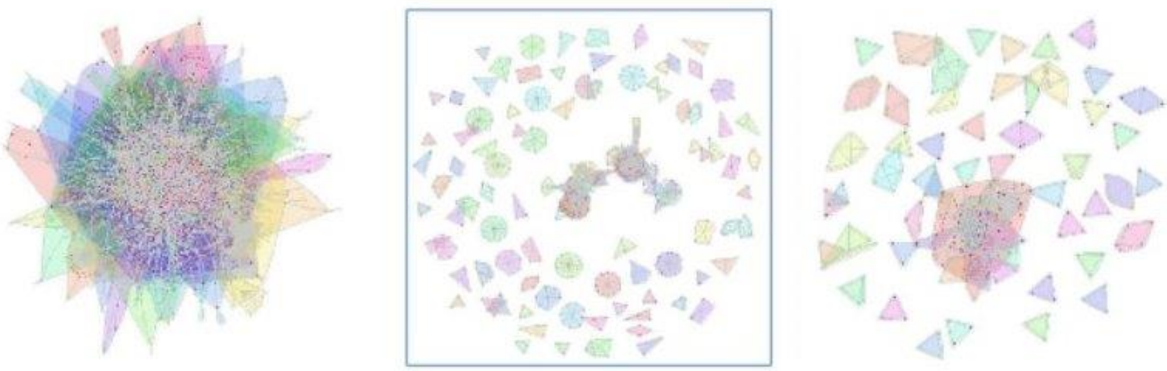
Figure 1 E L C similarity matrix and graph



IJSER

(a) (b)

Figure 2 L C similarity matrix and graph



(a) E L C (b) L C (c) C P M

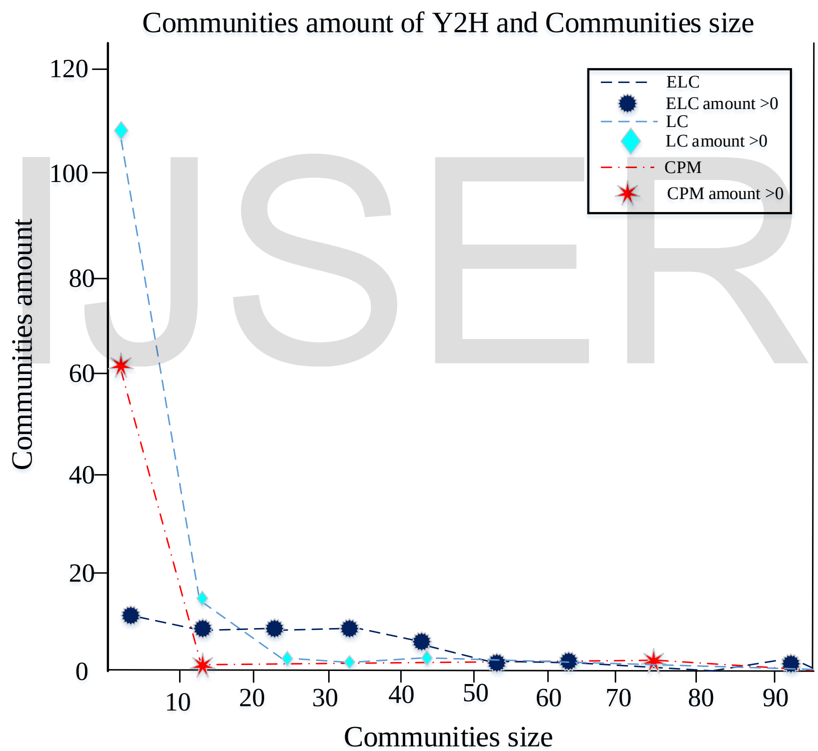
Figure 3 E L C, L C, C P M results in the diagram of Y 2 H dataset

Table 1 E L C, L C, C P M results in Y 2 H dataset

	I G P	P D	E Q
L C	0. 107	0. 065	0. 111
C P M	0. 077	0. 085	0. 062
E L C	0. 027	0. 005	0. 285*

Note: the data shown good value in similar place marked with (*).

As shown in figure 1 (a) and figure 2 (a), the L C method transform matrix less than the E L C. The E Q value is 0. 285 identified by the E L C method, the E Q value is 0. 111 identified by the LC method, and the CPM method identifier an E Q value is 0. 062. As shown in figure 4-15 (a, b and c), the inclusive of overall initial 1647 nodes and the E L C reached 54 groups. On other hand, the 958 nodes with 127 groups identified by the L C method. Then identified 63 groups by the C P M method.



IJSER

Figure 4) C P M, L C and E L C communities on nodes which Y2H statistics networks

From figure 4-16 as shown the C P M and L C have communities with higher p-value and less nodes, the E L C method obtains the number of nodes is high results than C P M and L C. In order to find out the communities of genes has statistical reasonable during whether the association, the Y 2 H dataset in groups of gene ontology enrichment identified by all the functional modules the 3 G O terms, these considered p-value lower than 0. 05, such as (C C, cellular component), (B P, biological process), and (M F, molecular functions) all 3 terms are shown in table 4-6, the lowest p-value=2. 96e-38 by C C and a cluster of 179 nodes obtain by the E L C, and the lowest p-value=1. 85e-17 by BP, and MF obtain p-value=1. 71e-14 with clusters of only 4 to 7 nodes obtain by LC method. The 3 methods all the p-value respectively C C, M F and B P the E L C has 5, 6 and 7 groups. We composed overall the p-values on 3 methods to researching the facts by use of mathematical the average classification measure found a solution on E L C are considerable large more than C P M, L C by G O groups at lowest p-value layer.

Table 2) Protein interactions Y 2 H result list

L C method						**C P M method**						**E L C method**					
B P		M F		C C		B P		M F		C C		B P		M F		C C	
G N	P value	G N	P value	G N	P value	G N	P value	G N	P value	G N	P value	G N	P value	G N	P value	G N	P value
7	1.85e-17*	4	1.71e-14*	44	5.11e-23	5	2.02e-13	4	2.16e-10	4	1.14e-12	179	5.30e-15	26	7.43e-13	179	2.96e-38*
5	2.02e-13	6	1.19e-11	4	1.63e-14	69	3.07e-10	5	2.34e-09	7	9.23e-12	35	6.14e-15	179	3.07e-10	20	2.50e-20
4	2.21e-11	4	2.16e-10	4	8.13e-14	3	5.02e-10	3	7.48e-09	3	2.52e-11	20	1.79e-15	9	5.90e-10	65	2.22e-15
44	1.95e-10	4	4.27e-09	20	5.23e-13	5	5.01e-09	3	5.76e-08	3	2.52e-10	65	6.64e-14	35	6.15e-10	26	6.27e-14
6	4.28e-10	3	5.76e-09	4	1.14e-12	3	5.52e-09	5	9.60e-07	5	2.52e-10	25	1.31e-13	45	7.68e-10	31	2.10e-13
3	5.02e-10	3	7.48e-09	6	1.87e-12	4	9.13e-09	7	1.12e-06	4	4.03e-10	96	3.76e-13	65	1.67e-09	31	1.25e-12
3	5.02e-10	44	1.64e-08	4	3.42e-12	5	3.27e-08	3	2.40e-06	5	4.46e-10	26	4.64e-13	156	1.80e-09	98	2.61e-12
4	7.50e-10	3	1.60e-07	6	5.17e-12	4	4.13e-08	4	3.20e06	5	1.21e-09	26	6.93e-13	18	5.78e-09	20	3.87e-12
8	1.52e-09	3	1.60e-07	7	9.23e-12	3	7.34e-08	3	4.48e-06	3	3.02e-09	373	4.13e-12	17	6.74e-09	156	5.82e-12
37	4.03e-09	8	2.48e-07	3	2.52e-11	7	7.36e-08	4	5.93e-06	4	1.13e-07	73	6.31e-12	8	5.12e-08	36	4.07e-11

Note: the data shown enrichment good value of G O marked with (*).

(1)

Experimental results of artificial datasets

Generate Strategy artificial of network

From communities artificially generated datasets are used here to compare the effects of E.L.C and each algorithm with an experimental defining similar dataset has 128 nodes cut into 4 groups and 32 each nodes. with respect to the degree the link desired average are definitions using the following generated and the proportion of group inside connected, the whole network be $\bar{n}$ allow to degree, and P_{inside} be the inside connects of group, while $P_{outside}$ is the proportion of outside connects during different group will be ($1-P_{inside}$), and then $P_{inside} \geq P_{outside}$ for appropriate groups. The relations nodes two chosen within every individual group combine the limitation procedure places connected sub-tree $\left[128*(\bar{n}/2)*P_{inside}\right]/4$. And the links outside nodes in different groups n it randomly the remaining $\left[128*(\bar{n}/2)*P_{outside}\right]$. As shown in

IJSER

experiments, the proportions $P_{\cdot inside}$ not only using different. However, a range of communities with situations different node also setting average degrees $\bar{n}$. Meanwhile, the node average degrees is 4, 8 and 12, showing 15 situations with locate of values and the proportion is modified form 0. .9 to 0. .5 by -0. .1 steps. The node degree grows and $P_{\cdot inside}$ drops, by expect that the overlapping during different groups increase disruptive. Overall, the effect values are the average over the 10 case by generated 10 communities. Whole artificially generated networks have 128 nodes and 4 known groups, and $128*(\bar{n}/2)$ connects with different artificial topologies.

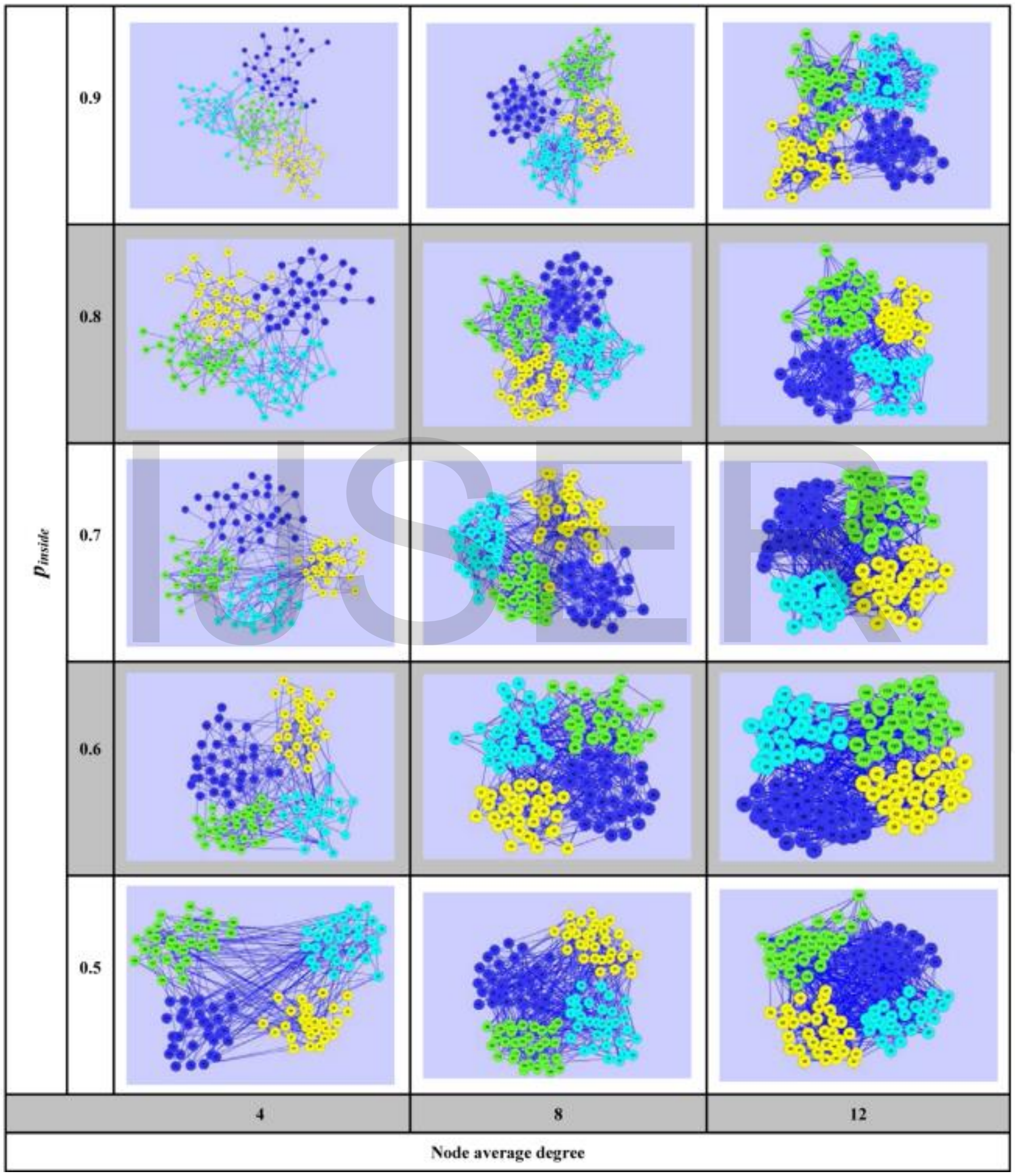


Figure Collection network structure diagram.

Analyze artificial of network

As shown in figure 4-17, that reveals a maximum degree consistent to more links during different community structure into the same in proportion to $P._{inside}$. Furthermore, while $P._{inside}$ is 0. .5. Whole groups are jointed and other disjoint group's outline not explicit at all. When proportion obtains 0. .9, the communities 4 each groups and tends to have lowest overlapping.

Table 1) E.L.C method results on the artificial dataset n

Degree average	12			8			4		
P inside	E Q	P D	I GP	E Q	P D	I GP	E Q	P D	I GP
0. .9	0. .205	0. .138	0. .329	0. .247	0. .138	0. .393*	0. .291*	0. .046	0. .347*
0. .8	0. .114	0. .135	0. .421	0. .151	0. .117	0. .395*	0. .227*	0. .034	0. .260*
0. .7	0. .059	0. .130	0. .318*	0. .093*	0. .082	0. .455*	0. .183*	0. .026	0. .227*
0. .6	0. .033*	0. .156*	0. .349	0. .070	0. .071	0. .375*	0. .173*	0. .024	0. .157*
0. .5	0. .026*	0. .204*	0. .420	0. .057	0. .065	0. .270*	0. .159*	0. .024	0. .164*

Note: the data shown good value in similar place marked with (*).

Table 2) L.C methods on the artificial dataset

Degree average	12			8			4		
P inside	E Q	P D	I GP	E Q	P D	I GP	E Q	P D	I GP
0. .9	0. .289*	0. .616*	0. .146	0. .295*	0. .839*	0. .224	0. .118	0. .319*	0. .056
0. .8	0. .177*	0. .586*	0. .098	0. .148	0. .452*	0. .071	0. .111	0. .436*	0. .096
0. .7	0. .090*	0. .557*	0. .118	0. .088	0. .152*	0. .109	0. .086	0. .327*	0. .046
0. .6	0. .016	0. .108	0. .725*	0. .078	0. .260*	0. .095	0. .086	0. .295*	0. .069
0. .5	0. .004	0. .114	0. .901*	0. .064	0. .208*	0. .095	0. .087	0. .284*	0. .061

Note: the data shown good value in similar place marked with (*).

Table 3) C.P.M methods results on the artificial dataset

Degree average	12	8	4

P inside	E Q	P D	I GP	E Q	P D	I GP	E Q	P D	I GP
0. .9	0. .201	0. .175	0. .640*	0. .274	0. .173	0. .162	0. .101	0. .221	0. .037
0. .8	0. .028	0. .090	0. .661*	0. .181*	0. .178	0. .099	0. .088	0. .189	0. .083
0. .7	0. .007	0. .094	0. .307	0. .091	0. .172	0. .103	0. .067	0. .151	0. .063
0. .6	0. .011	0. .107	0. .245	0. .087*	0. .210	0. .061	0. .066	0. .158	0. .077
0. .5	0. .016	0. .131	0. .221	0. .075*	0. .203	0. .118	0. .051	0. .120	0. .030

Note: the data shown good value in similar place marked with (*).

From table 1, that reveal a measure degree equal 4, irrespective of the value of $P._{inside}$, the extended link clustering method obtains the better average E.Q value more than each method. While the degree is 8, E.L.C value to take the consequence obtain the better value of E.Q at a proportion is 0. .7, and able obtain the better C.N value at each proportion. While the degree obtain 12, E.L.C able obtain the better value of EQ and values of P.D at the proportion is 0. .5 and 0. .6, for value of IGP, E.L.C often time has the best consequence, event while the degree obtain 12.

As shown in table 2, in addition to L.C method usually otain the better P.D value in addition to while the degree is 12 for a $P._{inside}$ is 0. .5, 0. .6, but one part of interest, thus it obtains the better E.Q values, while the average is 12, and while the $P._{inside}$ distance form 0. .7 reach 0. .9. the finally as shown in table 4-9. The C.P.M just obtains referable value for better degree values of EQ, while the degree average is 8.

As shown in table 1,2 and 3 that reveal the E.L.C method present to obtain more benefit with maximum $P._{inside}$, also maximum degrees, that is close the top left corner of the consisting as shown in figure 4-17. The possibly a value of the transform matrices it obtained. L.C possibly obtain best E.Q values with maximum P_{inside} and maximum degree, while be connected with the corner top right, when C.P.M able obtain appropriate divide with less value of P_{inside} referred by the less half of figure 4-17. That reveals L.C method only regard neighbor links, which C.P.M goals to look block for the maximum in the communities, shown by E.L.C regularly obtain the better I.G.P result, and regularly closest nodes in the homogeneous communities. That common the link clustering applied partition density value, it take the high PD value to cut the sub community with each small network.

The experimental results on the real world dataset can also reflect the conclusions obtained from the artificial datasets such as in the Y.2.H protein interaction network dataset, the karate union dataset and the dolphin dataset have less average degree, that is approximately respectively 3, 5 and 5, and consisting P_{inside} is near to 0. .9, 0. .8 and 0. .7 respectively, the whole area by E.L.C method. Otherwise, For the American politics book datasets has average degree approximately 9, and consisting P_{inside} close to 0. .7, that is on the overlapping area during E.L.C, and C.P.M methods have using comparison E.Q values. Although American University Football League dataset has average degree is approximately 10 and consisting P_{inside} is near 0. .6, that is only in the less half as shown in figure 4-17. So in this area, E.L.C, and L.C methods performs less more than C.P.M method.

IJSER

Summary

This papers has achieve conduct experiment base on the five real world network datasets, we propose calculate the link similarity, and bring into the more link information than the original link similarity method based on Jaccard distance, and improves the clustering analysis capability of the transformed matrix based on the improved link similarity calculation method, a new type of link clustering method extended link clustering is suggestion and the measure dataset of a lot of groups of networks achieve change for the better an overlapping community splitting up result more than link clustering and the original point clustering method ever obtained, which is closer to the real community result and also get the higher extended modularity EQ value.

The transformation matrix achieved by the similarity computation method offered in this paper be capable of clearly make a block matrix, but the amount of communities with the effect is more than the real partition result, which is based on hierarchical clustering and partition density. Therefore, how to find a more effective way to more accurately mining the matrix block in the similarity matrix will be the focus of the next step.

Future work and final words;

The key study content of this article is the application of natural computing in the detection of overlapping network communities. This paper researches in different factor such as class, cluster analysis, multi-objective optimization, single target optimization and overlapping community.

The further work, we will go on our study on interpreting the key sub structures and their capability significance of the complex network, as well as the social network, the key-word network, as well as the transportation traffic network, and focus on these structures' advancement to have a better considerate of the network's dynamicity.

The last, the network range of the algorithm in this paper is quite small, with a few tens and hundreds association of nodes. But in the real world, there are limitless networks of millions and millions relationship of nodes, the good organization of the algorithm is mainly important and very important, and progression algorithms are generally well suited for comparable processing, so here it is. At one point, it is worth further examination and study.

References:

[1]. Ellison N B, Vitak J, Gray R, et al. Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes.[J]. Journal of Computer-mediated Communication, 2014, 19(4):855–870.

[2]. Bayliss, D., & Villanustre, F. (2016). Graph Processing with Massive Datasets: A Kel Primer. Big Data Technologies and Applications. Springer International Publishing.

[3]. Hâncean M G. A Brief Overview of Social Network Analysis and its Current State within Romanian

Sociology[J]. International Review of Social Research, 2013, 3(3):5-11.

[4]. Sah P, Singh L O, Clauset A, et al. Exploring community structure in biological networks with random graphs[J]. Bmc Bioinformatics, 2014, 15(1):220.

[5]. Ji J, Lv J, Yang C, et al. Detecting Functional Modules based on a Multiple-grain Model in Large-scale Protein-Protein Interaction Networks.[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2016, 13(4):610-622.

[6]. Harakawa R, Ogawa T, Haseyama M. Accurate and Efficient Extraction of Hierarchical Structure of Web Communities for Web Video Retrieval[J]. Biochemical Journal, 2016, 4(1):49-59.

[7]. Chen X, Wolf T. Extensible hierarchical simulation of network systems[C]// Eighth Acm/ieee Symposium on Architectures for Networking & Communications Systems. IEEE, 2017:141-142.

[8]. Guy, M. (2016). The Open Education Working Group: Bringing People, Projects and Data Together. Open Data for Education. Springer International Publishing.

[9]. Santiago R, Lamb L C. Efficient quantitative heuristics for graph clustering[C]// the Genetic and Evolutionary Computation Conference Companion. 2017:117-118.

[10]. [Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., & Dawson, S. M. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behavioral Ecology & Sociobiology, 54(4), 396-405.

[11]. Enugala, R., Rajamani, L., Ali, K., & Kurapati, S. (2016). Identifying Natural Communities in Social Networks Using Modularity Coupled with Self Organizing Maps. Computational Intelligence in Data Mining—Volume 1. Springer India.

[12]. Yu, H., Braun, P., Yıldırım, M. A., Lemmens, I., Venkatesan, K., & Sahalie, J., et al. High-quality binary protein interaction map of the yeast interactome network. Science, 322(5898), 104-110.

[13]. He, J., Chen, D., & Sun, C. (2017). A fast simulated annealing strategy for community detection in complex networks. IEEE International Conference on Computer and Communications (pp.2380-2384). IEEE.

[14]. Prokhorenkova L O, Prałat P, Raigorodskii A. Modularity in several random graph models[J]. Electronic Notes in Discrete Mathematics, 2017, 61:947-953.

[15]. Prokhorenkova, L. O., Prałat, P., & Raigorodskii, A. (2016). Modularity of Complex Networks Models. International Workshop on Algorithms and Models for the Web-Graph (pp.115-126). Springer International Publishing.

[16]. Clancy, K. W., Weerapana, E., & Thompson, P. R. (2016). Detection and identification of protein citrullination in complex biological systems.Current Opinion in Chemical Biology, 30, 1-6.

[17]. Liu, H., & Wang, Z. (2017). A method of grouping user communication model based on clique percolation method. Electronic Technology.

[18]. Li M, Li D, Tang Y, et al. CytoCluster: A Cytoscape Plugin for Cluster Analysis and Visualization of Biological Networks.[J]. International Journal of Molecular Sciences, 2017, 18(9):1880

IJSER