ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology
An Indian Journal

FULL PAPER
BTAIJ, 10(13), 2014 [7310-7321]

Network community discovery algorithm based on multi-objective particle swarm optimization

Wan Li*, Yang Jie, Tang Pengfei
Department of Computer and Communication Engineering, Hunan University of Science and Engineering, Yongzhou, Hunan, 425199 (CHINA).

ABSTRACT

The quality of evaluation index has strong coupling correlation and data dependency in the evaluation of network community. To solve the problem of poor quality of traditional single evaluation compared with optimized network community discovery algorithm, this paper proposes online community discovery algorithm based on multi-objective particle swarm optimization. The algorithm generates Pareto optimal community classified collection through the optimization of multiple online community quality evaluation indicators at the same time, in which users can choose the most satisfied community structure according to the specific needs. Finally the comparison experiment is carried out between the single objective optimization method and multi-objective optimization algorithm. The experimental results show that the proposed algorithm can dig out higher quality online community in the absence of priori information and have higher stability of the system.

KEYWORDS

Online community; Particle encoding; Selection strategy; The objective function.

INTRODUCTION

Social networking, biological network, power network, Web and many other complex systems can be represented by a complex network, through two simple mapping, namely the mapping from the object to the network nodes and the map of relationship to the edge; a complex network can be represented as graph model. The research of complex networks is attracting the attention of scholars from physics, biology and computer science, sociology and other different fields. In addition to the well-known small-world and scale-free features, complex network also has the extremely important modularity, which implies rich community structures in complex networks model. According to the literature's description to Web communities, it can be loosely defined as the set of interconnected information carrier which has some common features, for example, collection of Web pages belong to a particular topic, micro group of weibo formed by the person who have a common interest, and so on[1-3]. From the point of network topology, an online community is a dense network diagram connected sub graph, between nodes within the subgraph, connection density higher than that of graph nodes internal and external connection density. The research achievements of a complex network community found has been successfully applied in areas such as terrorist organizations to identify, protein function prediction, public opinion analysis and processing, and many other fields.

Online community found complex network research is attracting wide attention of the researchers, methods have sprung up in recent years, Clauset A etc[4-6]. Carry out system analysis and the preliminary classification large number of methods, from the aspect of data mining, online communities found that the nature is based on the clustering study of network link, its goal is to network node set is divided into multiple internal links and external links closely sparse cluster. From the cluster learning perspective, the network community discovery algorithm's quality is largely depended on the quality of the network community structure design and optimization strategy of quality evaluation indicators.

At present, the most widely used community quality evaluation index is Q value function put forward by Girvan and Newman; it is defined through comparing the connection density of network graph and its corresponding random graph model of zero. The higher the Q value is, the greater the quality of the online community. However, the index exist disadvantage of 'particle size limitation'. Li z makes improvement for module density index Q Li. Evaluation indicators MinMaxCut trying to maximize node similarity in the community while minimizing the similarity between nodes in the community. MinMaxCut value is smaller, the higher the quality of the online community.

At present, the network community discovery algorithm in the objective function (network community structure quality evaluation indicators) optimization solution strategy can be roughly divided into two categories: basic heuristic method and heuristic method. The former transfer predefined complex network community discovery into designed problems of heuristic rules, according to the features of various community quality evaluation index design optimization strategy; Which use a variety of super heuristic operator in online communities found that the problem space quality evaluation indicators for optimization of the community[7].

Basic heuristics of community quality evaluation indicators can be divided into direct greedy method and indirect greedy method. Direct greedy method is very simple, is to initialize the network $|V|$ community, iterative process until the following algorithm termination conditions: calculating information gain degrees of the module relative to each side, to satisfy the largest community quality evaluation index increment edge to join, so as to realize the community merge. Representative algorithm of direct greedy method is module degree index Q optimization algorithm, the original Q optimization algorithm's time complexity is $O(n)(m+n)$ O (n) or $O(n_2)$.

In order to improve the efficiency and effectiveness of the algorithm, this paper puts forward a series of improvement methods: Clauset et al., designs the data structure Max - heaps will reduce time complexity to O (mdlogn), Danon put forward standardized processing to the Q value increment and found with the community has a bigger difference to the size of the community structure, Wakita put forward using the combined ratio (consolidationratio) on Q value incremental weighted to improve the scalability of the algorithm, Blondel put forward in the process of iteration merger just many community merger rather than just a merger between the two communities[8-9].

The basic idea of indirect greedy method is: the entire network is regarded as a community, and then loop the following process until the community quality evaluation index Q satisfies the given conditions: choose to have certain properties edge and remove to realize the online community. This method of selection strategy mainly includes: medium neutral (betweenness) supplied by priority, while clustering coefficient (clustering coefficient) for those priority, edge information center (information centrality) supplied by priority and edge stability factor (stability coefficient) supplied by priority. In addition to the greedy method was carried out on the side, Jin Wenyan put forward a kind of community discovery algorithm based on the topology potential, will each community as the topology potential field local high potential area, through the study of the greedy of local maximum potential value node optimization to realize the network community[10].

Basic heuristics thought simple and intuitive, easy to implement, but the method needs some prior knowledge defined recursive termination conditions, do not have ability of automatic identification the total network community, thanks in large part to limit such optimization method in the application of real complex network community discovery.

In order to overcome the lack of basic heuristics, researchers put forward a kind of super quality evaluation index to optimize community heuristic methods, mainly including optimization algorithm based on single objective and multi-objective. Tasgin and others realize the optimal approximation by using GA algorithm optimize community module degrees Q function[11]. Pizzutiz first give scores used to judge the quality of network partitioning community's (community score) definition, then use GA to optimize network partitioning-net[12]. Considering the massive amounts of social network, Lipczak put forward an ACGA algorithm based on the assumption of the community is small enough and several limited: a community coding will be for an individual, according to the community quality potentially increase amount to select individuals for genetic operation[13].

Duan Xiaodong et al. introduced particle swarm algorithm for implementing iterative binary network Web community discovery[14]. CDPSO algorithm uses particles encoding based on node neighbor table orderly, mining community by PSO global searching. Gog put forward a co-evolutionary algorithm based on individual information sharing mechanism for network community structure optimization, combined with GA variant algorithm CCGA of local search and the LGA, to achieve large-scale community discovery of complex networks through optimizing the community quality evaluation index Q. Zhu and Wang propose parallel genetic algorithm PGA of mining network.

Although these algorithm based on single objective optimization has better efficiency of time and can unearth meet network community structure of a specific goal, however, the network community founded problem of multiple targets in the actual application often need to balance, and these goals may be conflicting. Obviously, community discovery method based on single objective optimization can not meet the application requirements. Such as a result, community discovery based on multi-objective optimization began to concern. Gong Maoguo puts forward a kind of multi-objective optimization algorithm used for online community discovery which based on the mathematical programming method combined with evolutionary algorithm, and optimize the internal link density and the density of external links[15].

Multi-objective optimization algorithm (NNIA-Net[12], MOGA-Net, MOHSA and SCAH-MOHSA) are selected community score (community score) and community fitness (community fitness) as the optimization goal to achieve online community of digging, the difference is adopted by the super heuristic methods: NNIA -.net using immune algorithm, MOGA-Net with GA algorithm, adaptive hybrid multi-objective MOHSA harmony search algorithm, based on MOHSA SCAH-MOHSA adds a spectral clustering method.

Compared with the single objective optimization algorithm, the multi-objective optimization algorithm can carry on the comprehensive consideration of various community quality evaluation indexes; can found higher quality network community structure. Because of community discovery based on multi-objective optimization, there are some deficiencies, for example, the current algorithms are assumption indicators between network community quality evaluations may not be consistent, and not to assume that if I set up the theoretical proof or experimental verification, no research the relationship between properties of community quality evaluation index. In addition, the existing method of community discovery based on multi-objective optimization are almost based on genetic algorithm (based algorithm, GA), haven't seen the multi-objective particle swarm optimization algorithm for community discovery, and the related research has shown that multi-objective particle swarm optimization algorithm has excellent global optimization ability.

This article first from the Angle of the experiment verify the network quality evaluation indicators coupling relation and the existence of data dependency, which is deduced from the necessity of multi-objective optimization’ community discovery, then the problem of multi-objective optimization network community formal description, and then puts forward online community detection algorithm MOCD-PSO based on a multi-objective particle swarm optimization, the method through the optimization of multiple online community at the same time quality evaluation indicators generate Pareto optimal community classified collection, users can choose the most satisfied community structure according to the specific needs. The experimental results show that with the single objective optimization method (GN compared with GA-Net), or with multi-objective optimization algorithm (MOGA-Net and SCAH-MOHSA) compared with MOCD-PSO algorithm can be dug up higher quality of online community in the absence of a priori information.

NETWORK COMMUNITY STRUCTURE

Community can be classified into three categories: has the maximum density of internal links connected sub-graph, connected sub-graph with minimum density of the external links, at the same time has the largest internal link density and minimum density of the external links connected sub-graph. Researchers dig up different community’s definition from the network, and put forward several quantitative indicators to measure quality of online community. Community quality evaluation indicators (TABLE 1) are commonly used in this section on the basis of the given network community structure definition, TABLE 2 are symbols of various kinds of measures.

TABLE 1: Common community quality evaluation indicators

Metrics	Formula
Q	$Q=\sum_{i=1}^{m}\left[\frac{d^{lv}(v_i)}{d^{in}(v)}-\frac{d^{rule}(v_i)}{d^{in}(v)}\right]$
Q_{li}	$Q_{li}=\sum_{i=1}^{m}\left[\frac{d^{lv}(v_i)-d^{rule}(v_i)}{\lvert v_i\rvert}\right]$
MinMaxCut	$MMC=\sum_{i=1}^{m}\frac{d^{lv}(v_i)}{d^{rule}(v_i)}$

TABLE 2: Relevant symbols

Symbol	Remark
$d_j^{in} = \sum_{j*k/ev_j} A_k$	Internal degrees of node j in Community v_i
$d_j^{net} = \sum_{j*k/ev_j} A_k$	External degrees of Community v_i
$d^{lv}(v_i) = \sum_{j*k/ev_i} A_k$	Internal degrees of Community v_i
$d(v_i, v_j) = \sum_{j*k/ev_i} A_j$	Degree of association between Community v_i and Community v_j

Given the undirected network $G=(V,E)$, network node set v, network edge set as $E=\{E=(u,v)\,|u\in v,\ v\in v\}$, G with A size of $|v|x|v|$ of matrix A, if the side $E=(i,j)\in E$, $A_{ij}=1$; Otherwise, $A_{ij}=0$. Network community structure is a V m of network node set $p=(v_1,v_2,...,v_m)$, among them, the v_i must meet four conditions: $v_i \subseteq v$, $v_i \neq \varnothing (i=1,2,...m)$, $U_{i=1}^{m} v_i = v$ and $v_i \cap v_j = \varnothing (i \neq j)$.

COMMUNITY QUALITY EVALUATION INDICATORS

The community optimal quality evaluation index have maximization, has also minimize the optimal, both can be converted to. Do not break general, this article only talk about the optimal situation. The following, we first explore quality evaluation index of the coupling network community and the existence of data dependency through the experimental method, and then give the corresponding form definition.

Experiment 1 Checking community quality evaluation index's coupling correlation. The experimental data Karate network; Algorithm framework: CDPSO; Embedded algorithm framework community quality evaluation indexes: module Q, silhouette values GS.

Algorithm (CDPSO + Q) is a community quality evaluation index module Q for a single goal of network structure optimization algorithms, the algorithm (CDPSO + GS) is a community quality evaluation indicators silhouette values for a single target of network structure optimization algorithm. It can be seen from Figure 1 that the iteration number is 91, Q and GS Q = 0.415, $GS=0.790$ respectively; when the iteration number is 105, GS and Q, $Q=0.41$, $GS=0.777$ respectively. This shows that in the optimization module Q iteration process, the silhouette values are not synchronous optimization. Similarly, by the Figure 2 shows: in the iteration number is 85, GS and Q GS = 0.861, Q = 0.335respectively; When the number of iterations is 96, GS and Q $Q=0.133$, GS = 0.952respectively. This shows that in the iterative process of silhouette value, degree of module Q also didn't get synchronous optimization.

(a) (T=91,Q=0.415,GS=0.789)

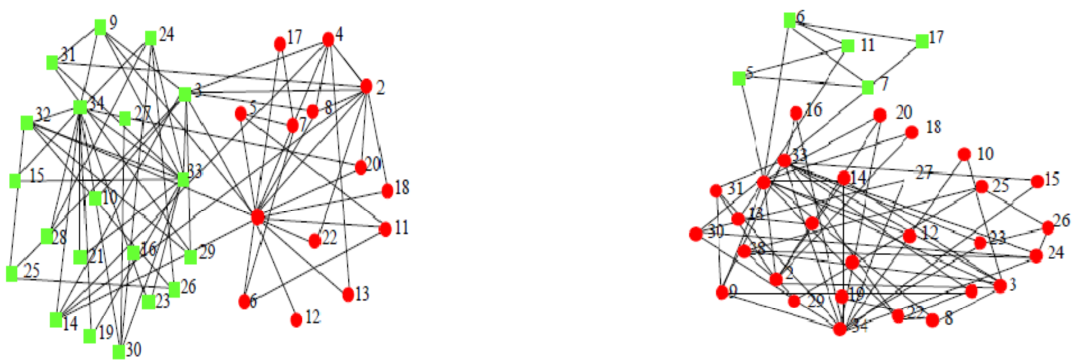
(b) (T=105,Q=0.42,GS=0.777)

Figure 1:Algorithm (CDPSO+Q) network community structure in the process of iteration

Experiment 2. Community quality evaluation index of data dependencies experiment. Experimental data: data generator of the LFR network structure of six artificial network SynNet_1 ~ SynNet_6; Algorithm framework: CDPSO; Algorithm embedded within the framework of the community of quality evaluation indicators: module Q, GS silhouette values. By using different evaluation index optimization strategy of network community structure of F - Measure value shown in TABLE 3. F - Measure values of The F - Measure definition, network community structure, the greater indicates are the higher quality will be. See from TABLE 3, GS module degree, compared with Q as the optimization goal of the algorithm can find better SynNet_2, SynNet_3, SynNet_4 implied in SynNet_6 community structure, similarly, module Q, compared with index as the optimization goal of GS algorithm can unearth SynNet_1 with better quality of SynNet_5 community structure. The experiment shows two different evaluation indexes in optimization of different network data set is not the same performance.

Network community quality evaluation indicators coupling relationship with the form of data dependencies are defined as follows:

Given network $G = (V, E)$, $|v| = N|$, community partition function clusterer: $v \to p, |p| = 2N, F =$, $(F_1, F_2, ..., F_{K1} - F_K)$ the community quality evaluation index, among them, the F_i: $p \succ R(p \in p, I = 1, ...k)$.



(a) (T=85,GS=0.861,Q=0.335) (b) (T=96,Q=0.133,GS=0.952)

Figure 2: Algorithm (CDPSO + GS) network community structures in the process of iteration

Nature 1 (Correlation coupling). For network G, in the community partition algorithm clusterer the t_m with the t_n iteration $(m \succ n)$, community quality evaluation indicators are F_i $F_i(tm) \succ F_i(t_n)$, $F_j(t_m) \prec F_j \wedge j \neq i$ Sunday afternoon j indicates I, said the F_i and F_j coupling relationship exists.

TABLE 3: Data dependencies of modules Q and GS

Dataset	(CDPSO+Q)	CD[SP+GS
SynNet_1	0.73	0.74
SynNet_2	0.74	0.69
SynNet_3	0.73	0.71
SynNet_4	0.71	0.73
SynNet_5	0.72	0.76
SynNet_6	0.74	0.71

Property2. (Data dependencies). For network G1, the prior to community structure p_1^{true}, clustering validity function CV (such as F - Measure, Clustering Error), community partition function clusters respectively by using optimization evaluation index F_i and F_j to divide network G1, with community structure p_2^i and p_1^j were obtained, the F- Measure $(p_1^{true}, p_1^i) \succ F$ - Measure(p_1^{true}, p_1^j), remember to:

$$V = Y \cup h \tag{1}$$

And with a priori figure G_2 community structure p_2^{true} of the network

$$E = \{e(y_i, h_j) | y_i \in y, h_j \in h, k_{ij} \succ 0\} \tag{2}$$

Community quality evaluation index F_i and F_j have strong data dependencies.

Theory, according to a priori weights to different evaluation index synthetic evaluation index, and then USES the single objective optimization strategy. But because of the existence of nature 1, namely the coupling relation between each evaluation index, makes it difficult to ensure the optimal solution.

Can be seen from the experiment 2, for a specific network, does exist an optimal evaluation indicator, but the optimal evaluation index is varied from the network, and cannot know in advance. Therefore, the existence of the nature 2 shows only select some quality evaluation index for network community was found has limitations.

MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION ALGORITHM

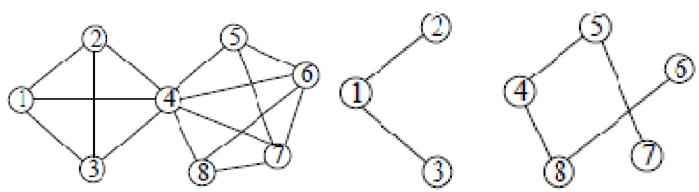
Particle coding

Particle coding using coding method based on node neighbor table orderly in the CDPSO, its basic idea is: first of all nodes in the figure number, then the neighbors of each node Numbers sorted according to its neighbor table orderly formation, position update phase particles in the initialization or generate new particles, ensure the legitimacy of the individual. In Figures 3 (a) in the network as an example, the first to establish neighbor ordered lists of all nodes (FIG. 3 (d)), according to this table can network community structure (FIG. 3 (b)) for individual coding, the result is shown in Figure 3 (c). The encoding process is simpler, the following to node 1 in 3 (b) as an example to illustrate: for nodes in Figure 3 (b) 1, linked with the node 2, Figure 3 (d) the central node 1 neighbor ordered list, the nodes 2 is the first element in the ordered list, therefore, in Figure 3 (c), the particle's first Dim $Dim(=1)d$ has a value of 1; Similarly, you can get corresponding values by other Dim particles. The encoding method has the following three advantages:

1) Avoid illegal particles produced, which can reduce the number of constraint conditions in a multi-objective optimization problem.

2) Automatically determine the number of community;

3) avoid iteration 2 of the binary coding based on the strategy of easy to fall into local optimum situation.



Dim	1	2	3	4	5	6	7	8
Pos	1	1	1	3	4	2	1	1

Figure 3: Particle number of CDPSO

Objective function

MOCD - PSO algorithm chooses the most common three community quality evaluation indicators (Q, MaxMinCut and GS) as objective function. To construct multi-objective optimization problem of network community structure (formula (3) to formula (4).

$$y' = \left\{ y_i \middle| y_i \in y, des(y_i) \succ \lambda \right\} \tag{3}$$

$$y' = \underset{i=1}{\overset{n}{y}} c_i,\ c_i \neq \varnothing, and\ i \neq j, c_i \cap c_j = \varnothing \tag{4}$$

Particle update policy

Particle swarm optimization (Pos) algorithm is put forward, mainly inspired in the biology research results: birds in foraging, memory and the optimal foraging path, sharing what they found by this sharing information, the birds can find better food faster. The algorithm in solving optimization problems, will every solution of optimization problem corresponds to the search space of a bird, or called "particles", all the particles can be calculated by the optimization function; At the same time, each particle has the corresponding fly speed determines the direction and distance. That is to say, the whole optimization process is based on the particle's position and speed continuously updated.

The particle's velocity has a lot of updating method, basically can be summarized as formula (5) as follows:

$$E_{s_j}(c_i) = \left\{ y \middle| y \in c_i\ and\ (y, s_j) \in R \right\} \tag{5}$$

$$Amb_{ij} = \begin{cases} \left| c_i - q_{s_j}(c_i) \right|, q_{ij} \geq \sigma \\ \left| q_{s_j}(c_i) \right|, q_{ij} \geq \sigma \end{cases} \tag{6}$$

Among them, $x_i = \left(x_{i1}, x_{i2}, ...x_{id}\right)$ and $x_{ij} = \left(x_{ij1}, x_{ij2}, ...x_{ijd}\right)$ represent position and speed of particle I, t for the evolution algebra, w as the inertia coefficient, c1 and c2 for learning factor, rand $(.)$ is the random number uniform distribution on [0, 1], the $p_i = \left(p_{i1}, p_{i2}, ...p_{id}\right)$ is the optimal position, the history optimal location of particle I is $p_g = \left(p_{g1}, p_{g2}, ...p_{gd}\right)$, $p_g = \left(p_{g1}, p_{g2}, ...p_{gd}\right)$ is current field of the optimal particle location. The role of the formula (6) is to the particles in the solution space flight Pi speed adjustment, namely, through memory itself searched the optimal position of the $p_i = \left(p_{i1}, p_{i2}, ...p_{id}\right)$ so far to realize the particles from learning and sense the whole particle swarm search to the optimal position of the $p_g = \left(p_{g1} \; and \; p_{g2}, ...p_{gd}\right)$ so far to implement the group information sharing.

Because p_g is with the best fitness particles in the particle swarm, also known as the leader, the Pleader, its main function is to guide the optimal solution of the particle swarm to contain potential solution area direction. Particles neighborhood topology determines the identity of the Pleader. For example: when the particles neighborhood topology is circular, the $Pleader = Plbest$; When the particles neighborhood topology is full connection, $Pleader = Pgbest$. the particles neighborhood topology is star, $Pleader = Pfocal$. MOCD - PSO USES the whole connection of particle neighborhood topology structure.

The particle position update strategy of traditional discrete PSO can form into formula (4), MOCD - PSO to improve for formula (5), formula (6) as follows:

$$Amb_{ij} = \sum_{s_j \in s} Amb_{ij} \tag{7}$$

$$Amb = \log\left[\frac{\sum_{c_i \in Clus} Amb_i}{|Clus|}\right] \tag{8}$$

Among them, the k is in addition to the current connection neighbor any random neighbor nodes, namely $k = ceil\left(rand \times \deg\left(v_j\right)\right) \wp k \neq x_{ij}\left(t\right)$; On the ceil of integral function; $\deg\left(v_j\right)$ represents node v_j degrees (and node in the graph G, v_j associated number of edges); p for a predetermined threshold. The meaning of the formula (4) can be simply described as follows: through s-shaped function to map of particle velocity, if the value is greater than the predetermined threshold mapping, the i_{th} a weight assignment of this particle positions vector as different values for the particle and the current component of neighbor nodes. Based on the location update formula (5) of the particle coding method in section A can make the particle has more strong searching ability, and the formula (6) is a s-shaped function $sig\left(v_{ij}\left(t+1\right)\right) = 1/\left[1 + \exp\left(v_{ij}\right)\left(t+1\right)\right]$, its purpose is to improve the convergence of the algorithm, at the same time, limit v_{ij} value within the interval $\left[5,5\right]$, in order to prevent the sig saturation function.

Leader selection strategy

As multi-objective discrete PSO algorithm, MOCD - PSO will be formed in the process of iterative optimization of multiple non dominated solutions (non-dominated), which can not meet the needs network partitioning scheme collection Non Set of Pareto dominance relationship, this proposed how a leader in particle update to make a choice. MOCD - PSO based on kernel density estimation leader selection mechanism for leader selection (Figure 4). To simplify the instructions, make $|objectives| = 2$, $|Non\,Set| = 10$ |, that is, to optimize the target number is 2, there're a total of 10 kinds of algorithms for the current iteration process can not meet the network partitioning scheme needs of Pareto dominance relationship, distributed in the plane composed of goals 1 and 2.

For each point x in the Non Set, there is a r neighborhood Neighbor $\left(x,r\right)$ to that point as the center, calculate average distance dist from the center point to other points in its neighborhood, choose the largest average leaders as the center. If there is many such a leaders, choose the most r neighbors for leaders; if the leaders contain more than one leader, to a randomly selected from it. This process can form into algorithm leader Selection. If the particle r neighborhood neighbor number N_b , has the time complexity of the algorithm is $o\left(N_b \; x\left|Non\,Set\right|\right)$.

Algorithm: Leader Selection.

Input: meet the dominance in online community plan collection $Non\,Set$, neighborhood radius r .

Output: particle flight leader.

Step 1 $leaders = \arg\max_{x \in NonSet} \frac{\sum_{j \in Neighter} dist(x,y)}{|Neighbor(x,r)|}$;

Step 2 if $|leaders| \succ 1$, so $leaders = \arg\max_{x \in NonSet} |Neighbor(x,r)|$;

Step 3 if $|leaders| \succ 1$, so $leaders = randomSelec(leaders)$

Step 4 return leaders.

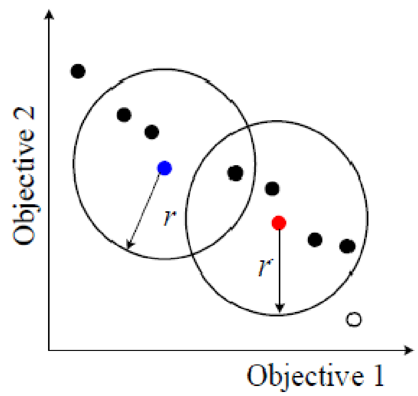


Figure 4: Leader selection mechanism based on kernel density estimation

Algorithm description and analysis

The MOCD - PSO algorithm described as follows:

Step 1 Set the particle swarm size, scope and the particle position, dimension and velocity, particle swarm inertial factor, neighborhood radius, and the largest external archive Capacity;

Step 2 Calculate the particle fitness vector $F(p)$;

Step 3 Comparing particle Pareto dominance relations;

Step 4 Use Update PS algorithm to update the Pareto optimal set community structure;

Step 5 To establish a network neighbor node number list of all nodes;

Step 6 The coding method based on node orderly neighbor table to initialize particle swarm;

Step 7 Use leader Selection algorithm choose particle flight leader.

Step 8 According to the formula (7), formula (8) to update the particle's position and speed;

Step 9 Repeat Step 4 to Step 7, until the number iterative algorithm to achieve the user to specify the maximum numerical iteration time, output all the Pareto optimal solution set elements of the network community structure.

MOCD - PSO algorithm can be divided into two large Step: the first piece is Step 1 to Step 3, mainly be responsible for the relevant parameters and data structure needed by perform initialization algorithm; the second piece is Step 4 to Step 9, realize the multiple objective function optimization mainly through particle swarm collaboration in the solution space flight, if meet the end of the iteration optimization conditions, the output Pareto optimal solutions corresponding to the online community, end of the algorithm.

EXPERIMENTAL ANALYSIS

Data set

TABLE 4: Experimental network

Networks	Nodes	Edges
Karate	35	77
Dolphins	63	160
HLM	78	120
SynNet_1	103	274
SynNet_2	153	450
SynNet_3	200	525

We experiment analysis was carried out on the MOCD - PSO algorithm in 3 real network and 3 artificial on the network. 3 real network respectively is: the first is to help the network, the network is in an American university social relationship between Karate club members, a total of 34 node, 78 link edge; The second is the Dolphins network, a total of 62 nodes, 159; The third is HLM network, the network contains 77 nodes, 121, mainly based on the kinship in the classic a dream of red mansions media relations with typical structure of five family (Ning guo mansion, history of Rong guo mansion, Wang fu and Xue Fu) between the networks of social relationship. Three artificial networks is simulation generated by the LFR data generator network. The characteristics descriptions of the experimental network are shown in TABLE 4.

Convergence analysis

To evaluate convergence of MOCD-PSO algorithm, we introduce the generation distance (generation short, GD) as evaluation standard, GD index is defined as the formula (9),(10). The smaller the GD, shows that the obtained Pareto optimal solution set, the approximate global optimal solution set, and the algorithm convergence is also better.

$$ca_{ij} = \begin{cases} \dfrac{\sum\limits_{y_k \in c_i} a_{kj}}{|c_i|}, & q_{ij} \geq \sigma \\ 0, & q_{ij} \prec \sigma \end{cases} \tag{9}$$

$$diff\left(c_i, c_j\right) = \frac{cv_i . cv_j}{|cv_i| . |cv_j|} \tag{10}$$

Among them, | PS | for the number of the Pareto optimal solution concentration solution, f_i is the minimum Euclidean distance of Pareto optimal solution to the global optimal solution. MOCD-PSO algorithm is presented in TABLE 5 the income of GD index respectively the average (mean) and variance (var) for six experimental network running 50 times. Can be seen from TABLE 5, no matter in a real network, or on the artificial network, MOCD-PSO algorithm shows a better convergence.

TABLE 5: GD indicators on the different test network

Tested Networks	Mean	v_{ar}
Karate	1.2634e-3	2.7632e-7
Dolphins	5.5432e-3	3.4621e-7
HLM	1.0321e-2	6.3413e-5
SynNet_1	4.2412e-2	6.5532e-5
SynNet_2	4.8322e-2	3.9321e-5
SynNet_3	6.9421e-2	7.4401e-5

This section from distribution uniformity and dispersion two aspects of the Pareto optimal solution, the generated Pareto optimal community structure set are analyzed through MOCD - PSO algorithm. The first SP index is used to describe the Pareto optimal solution in the distribution uniformity of target space, make the network node number is r, represents the k d of the I particle f_i^k, I SP indicators may be defined as formula (11) to formula (12), the smaller the SP, means that the more uniform Pareto optimal solution set distribution.

$$dvst = \frac{\sum\limits_{c_i \in Clus, c_j \in Clus} diff\left(c_i, c_j\right)}{|Clus|} \tag{11}$$

$$a = \frac{q\left(r|R\right)}{q\left(r\right)} = \frac{q\left(r\right).q\left(R|r\right)}{q\left(R\right).q\left(r\right)} = \frac{q\left(R|r\right)}{q\left(R\right)} \tag{12}$$

We put the Pareto optimal community structure set's traditional update strategy compared with heuristic update strategy of this paper, the MOCD - PSO algorithm is presented in TABLE 6 the income of the SP index respectively the average (mean) and variance (var) for 6 experimental network running 50 times.

As can be seen from TABLE 6, heuristic update strategy of this paper can make the solution of Pareto optimal community structure set distribution more uniform. Then use the Pareto optimal solution set of the statistic range (range) to analyze its dispersion (formula (13)), can be seen from TABLE 7, heuristic update strategy of this paper can make the solution of Pareto optimal community structure set distribution more dispersed, and the number of the optimal solution is related to the number of network nodes and the network related modules.

TABLE 6: Pareto distribution uniformity of optimal community structure set

Tested Networks	Traditional approach		Our approach	
	Mean	v_{ar}	Mean	v_{ar}
Karate	3.2644e-3	4.9632e-4	1.8634e-5	3.8634e-7
Dolphins	6.8435e-3	1.2625e-4	2.7432e-5	1.6621e-7
HLM	5.8322e-3	2.5417e-4	8.9321e-4	5.0412e-5
SynNet_1	2.5418e-3	1.5031e-3	2.2416e-4	1.9534e-4
SynNet_2	4.7323e-3	3.5326e-4	2.8323e-3	3.1331e-4
SynNet_3	3.3421e-3	3.1402e-3	2.6427e-2	1.6492e-5

$$MAP = \frac{\sum_{k=1}^{y} AP(k)}{Y} \quad (13)$$

TABLE 7: Pareto dispersion of optimal community structure set

Tested networks	Traditional approach				Our approach			
	# of	Width of the range			# of	Width of the range		
	Solutions	$F_1(p)$	$F_2(P)$	$F_3(P)$	Solutions	$F_1(P)$	$F_2(P)$	$F_3(P)$
Karate	11	0.3867	0.3211	0.4021	21	0.3884	0.3521	0.4782
Dolphins	16	0.4033	0.3676	0.5432	35	0.4236	0.3780	0.6342
HLM	57	0.5762	0.4872	0.5873	86	0.5783	0.4863	0.6031
SynNet_1	99	0.6489	0.5210	0.7534	100	0.6842	0.5123	0.815
SynNet_2	100	0.6231	0.5372	0.7467	100	0.6412	0.5563	0.7801
SynNet_3	100	0.5832	0.5321	0.6898	100	0.6034	0.5321	0.7142

Effectiveness analysis

In order to evaluate the effectiveness of the MOCD-PSO algorithm, we use the normalized mutual information (normalized mutual information, NMI) evaluation criterion to measure MOCD-PSO algorithm's calculation result is consistent with the real network community (formula (13)), the greater the degree of the value, the means that the higher the degree of match the calculation results and the network real communities. The MOCD-PSO algorithm compare and analyze with four representative community discovery algorithm (GN, GA-Net, MOGA net and SCAH-MOHSA), among them, the former both for single objective optimization algorithm, the after for multi-objective optimization algorithm. All kinds of algorithm is presented in TABLE 8 the income of the mean and variance of NMI in 6 experimental respectively on the network running 50 times.

$$y' = \underset{i=1}{\overset{n}{y}} c_i,\ c_i \neq \varnothing, and\ i \neq j, c_i \cap c_j = \varnothing \quad (14)$$

Can be seen from TABLE 8, with single objective algorithm GN compared with GA-Net, the quality of the community structure is found by the MOCD-PSO algorithm is far higher than the quality of the results of both the community structure, in the community, such as network with strong modularity structure Karate and *Dophins*, can accurately identify the network community structure. MOGA-Net with multi-objective optimization algorithms compared with SCAH-MOHSA, not only from the average Min_Avg Min (formula (14)) sense, MOCD show a clear advantage PSO algorithm, and the variance Max_Std (formula (15)) sense, MOCD - PSO algorithm is the best results of community structure of NMI distribution more uniform, therefore, our method has stronger robustness.

$$diff\left(c_i, c_j\right) = \frac{cv_i . cv_j}{\left|cv_i\right| . \left|cv_j\right|} \tag{15}$$

$$ca_{ij} = \begin{cases} \dfrac{\sum_{y_k \in c_i} a_{kj}}{\left|c_i\right|}, & q_{ij} \geq \sigma \\ 0, & q_{ij} \prec \sigma \end{cases} \tag{16}$$

TABLE 8: Network community quality comparison

Tested Networks	GN	GA-Net	MOGA-Net		MOCD-PSO	
			Max _ Avg	Max _ Std	Max _ Avg	Max _ Std
Karate	0.623	0.864	1	0	1	0
Dolphins	0.571	0.881	1	0	1	0
HLM	0.651	0.852	0.932	0.037	0.938	0.027
SynNet_1	0.591	0.779	0.815	0.046	0.936	0,027
SynNet_2	0.604	0.814	0.853	0.048	0.841	0.035
SynNet_3	0.651	0.856	0.877	0.051	0.887	0.033

ca_{ij} represents the first res results community structure generate by compare algorithm; ca_{ij} represents network real community structure; v_m represents the first m community in ca_{ij}; v_n represents the n community in ca_{ij}; ca_{ij} represents algorithm is calculated in the first run time experiment set network community structure; NMI_{run} represents the run time of all network community structure experiment and the maximum of NMI; Here RUNS represents the number of experiment, values of 50 ($runs = 50$).

CONCLUSION

Discovery and research of complex network community is great significant in many aspects of the Internet culture security and information personalized service, at present, most of the complex network community discovery algorithm based on biological inspired optimization is based on a single measure target to drive the quality of the community structure and the quality evaluation index of the community diversity makes network community structure analysis of the practice of staff in the face of numerous evaluation index is difficult to decision-making, and reveal the real community structure of complex networks is not reality to a single measure can be fully reflected.

In this paper, through a large number of experiments found data dependency and coupling relationship of community quality evaluation index, analyzes these two kinds of nature lead to network community discovery algorithm based on a single evaluation index optimization has limitations. Therefore, this article found that the problem further make complicated network community formal to multi-goal optimization problem, at the same time network community detection algorithm MOCD-PSO is proposed based on multi-objective particle swarm optimization, in this algorithm, we put forward the new Pareto optimal network community structure set update strategy and implementation ably, selecting module index Q, minimum max-cut index MinMaxCut and silhouette (silhouette) index to build optimization of MOCD-PSO.

Experimental results show that the MOCD-PSO algorithm has good convergence, and can discover uniform distribution and Pareto optimal network structure set of higher dispersion, and compared with the single objective optimization method, or with multi-objective optimization algorithm (MOGA-Net and SCAH-MOHSA), Pareto optimal network structure and the community structure within network is more consistent is found by MOCD-PSO algorithm.

ACKNOWLEDGMENT

The paper is supported by key discipline for computer application and technology of Hunan University of Science and Engineering. This work was supported in part by a grant from the Research Foundation of Education Bureau of Hunan Province, China (Grant No.12C0681), and supported in part by a grant from the Research Foundation of Education Bureau of Hunan Province, China (Grant No.13C336), and supported in part by a grant from the Science and Technology Projects for Yongzhou City in 2013(Grant No. [2013]17).

REFERENCES

[1] Wonki Kim, Jasung Koo, Jechang Jeong; "Fine Directional Interpolation for Spatial Error Concealment", IEEE Transactions on Consumer Electronics , **52(3)**, 1050-1056 **(2006).**

[2] Y.Zhao, D.Tian, M.M.Hannukasela, M.Gabbouj; Spatial Error concealment Based on Directional Decision and Intra Prediction, IEEE International Symposium on Circuits and Systems, **3**, 2899-2902 **(2005)**.

[3] Nhon.H.Trinh, Benjamin.B.Kimia; "Skeleton Search: Category-Specific Object Recognition and Segmentation Using a Skeletal Shape Model", International Journal of Computer Vision, **94(2)**, 215-240 **(2011)**.

[4] Muhammad.J.Mirza, Nadeem Anjum; Association of Moving Objects Across Visual Sensor Networks; Journal of Multimedia, **7(1)**, 2-8 **(2012)**.

[5] J.He, Y.Geng, K.Pahlavan; Modeling Indoor TOA Ranging Error for Body Mounted Sensors, 2012 IEEE 23nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Sydney, Australia, 682-686, Sep.**(2012)**.

[6] Chen Jie，Xin Bin，Peng Zhihong; Statistical learning makes the hybridization of particle swarm and differential evolution more efficient-A novel hybrid optimizer. IEEE Transactions on Evolutionary Computation, **6(3)**, 239-251 **(2008)**.

[7] Yu Chien-Chih，Wang Chen-Shu; A hybrid mining approach for optimizing returns policies in e-retailing; Expert Systems with Applications, **35(4)**, 1575-1582 **(2008)**.

[8] K.Wang, H.Xie; "Texture image segmentation using fully global minimization active contour model, ", 2011 International Conference on Multimedia Technology (ICMT), 2649-2653, 26-28, July **(2011)**.

[9] S.Li, Y.Geng, J.He, K.Pahlavan; Analysis of Three-dimensional Maximum Likelihood Algorithm for Capsule Endoscopy Localization, 2012 5th International Conference on Biomedical Engineering and Informatics (BMEI), Chongqing, China ,721-725, Oct. **(2012)**.

[10] Geyer R，Van Wassenhove L N，Atasu A. The economics of remanufacturing under limited component durability and finite product life cycles. Management Science, **53(1)**,88-100 **(2007)**.

[11] Gong Maoguo，Jiao Licheng，Liu Fang，et al.Memetic computation based on regulation between neural and immune systems：the framework and a case study.Science China：Information Sciences,**45(11)**, 2131-2138 **(2010)**.

[12] Y.Geng, J.He, H.Deng, K.Pahlavan; Modeling the Effect of Human Body on TOA Ranging for Indoor Human Tracking with Wrist Mounted Sensor, 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, Jun. **(2013)**.

[13] A.Beristain, M.Grana, A.I.Gonzalez; "A Pruning Algorithm for Stable Voronoi Skeletons", Journal of Mathematical Imaging and Vision, **42(2-3)**, 225-237, **(2012)**.

[14] H.B.Ren, Y.X.Zhu, G.Y.Xu etc; Vision-Based Recognition of Hand Gestures: A Survey. Acta Electronica Sinica, **2(28),** 118-122 **(2000)**.

[15] Z.W.Xu, M.L.Zhu; Optimum Colorspace for Skin-tone Detection; Journal Of Computer-Aided Design & Computer Graphics, **18(9)**, 1350-1356 **(2006)**.