SCIENTIFIC REP🅞RTS

OPEN

Complex Network Clustering by a Multi-objective Evolutionary Algorithm Based on Decomposition and Membrane Structure

Ying Ju[1], Songming Zhang[1], Ningxiang Ding[1], Xiangxiang Zeng[1] & Xingyi Zhang[2]

The field of complex network clustering is gaining considerable attention in recent years. In this study, a multi-objective evolutionary algorithm based on membranes is proposed to solve the network clustering problem. Population are divided into different membrane structures on average. The evolutionary algorithm is carried out in the membrane structures. The population are eliminated by the vector of membranes. In the proposed method, two evaluation objectives termed as Kernel J-means and Ratio Cut are to be minimized. Extensive experimental studies comparison with state-of-the-art algorithms proves that the proposed algorithm is effective and promising.

Network analysis is an important topic in computer science and bioinformatics researches[1–3]. In the last decade, a large number of algorithms for network clustering, such as the algorithm of Girvan and Newman[4], fast greedy modularity optimization[5], and Markov cluster algorithm[6], were proposed. Much research, such as that by Orman and Labatut[7] and Forunato[8], has been conducted on community detection in networks. However, because of the inherent complexity of network clustering, community detection problems often cannot be satisfactorily solved by traditional heuristic optimization methods. Thus, metaheuristic algorithms are adopted as a tool for dealing with community detection problems. These metaheuristic algorithms are notable for their effective local learning and global searching capabilities and have proven to be more successful than heuristic methods in solving optimization problems[9–12]. Metaheuristic algorithms, which also automatically determine the number of communities, are thus convenient to adopt in community detection applications[13]. Numerous scholars have applied some metaheuristics algorithms, such as evolutionary and particle swarm optimization algorithms, to the problem of network clustering. In this regard, Pizzuti[14] proposed a single objective genetic algorithm (GA-net) for network clustering; Gong[15] suggested a memetic algorithm-based network clustering method (Memenet). Although used successfully, in both theory and application in community detection problems, those single objective algorithms still have a significant disadvantage. For example, different algorithms on the same network may produce different solutions. Many single objective algorithms also must determine the number of communities as prior information[16,17]. However, for real networks, this information is often unknown.

To alleviate the disadvantages of single objective algorithms, multi-objective optimization is applied to the problem. A large number of multi-objective optimization evolutionary algorithms have been developed, which can be potentially effective and helpful for solving the problem[18–20]. Therefore, using multi-objective optimization algorithms to solve the community detection problem has become a significant subject[21,22]. In 2007, Zhang and Li[19–21] proposed an algorithm, a Multi-Objective Evolutionary Algorithm based on Decomposition MOEA/D). However, research on MOEA/D has revealed that some, but not all, solutions are chosen in several sub-problems (Fig. 1), which may result in loss of population diversity. To address the sub-problems, we designed a new Multi-Objective Evolutionary Algorithm, where each sub-problem has several solutions (Fig. 2).

Membranes play an essential role in the structure and the functioning of living cells. In this study, we propose a novel algorithm, Multi-Objective Evolutionary Algorithm based on Decomposition and Membrane structure (MOEA/DM), for community detection based on an evolutionary algorithm. To solve the problem of a solution corresponding to several sub-problems, we add a membrane structure to help ensure that a sub-problem will have multiple solutions, where the membrane structure refers to the structure of membrane computing models[23–26].

[1]School of Information Science and Technology, Xiamen University, Xiamen, China. [2]School of Computer Science and Technology, Anhui University, Anhui, China. Correspondence and requests for materials should be addressed to Xx.Z. (email: xzeng@xmu.edu.cn) or Xy.Z. (email: xyzhanghust@gmail.com)
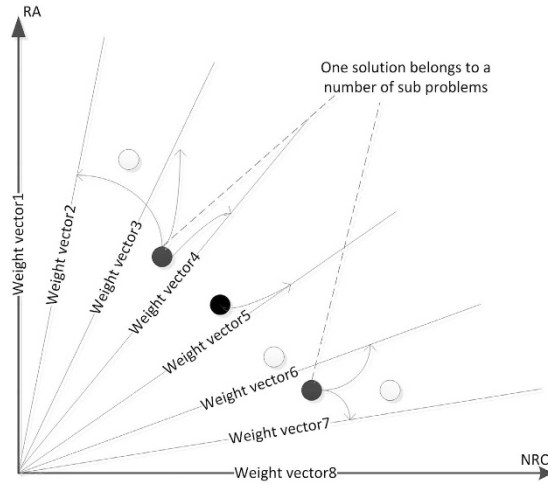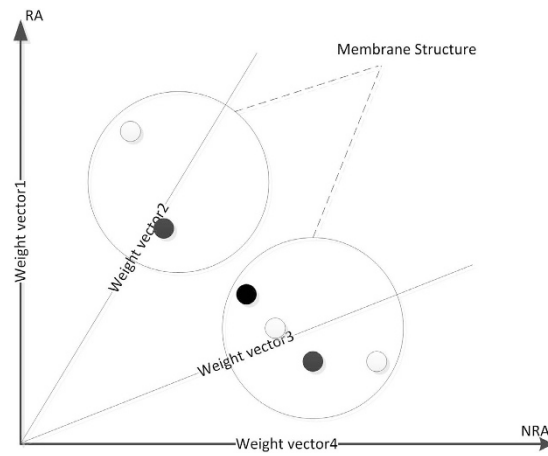
Figure 1. Solution structure of MOEA/D algorithm.

Figure 2. Solution structure of MOEA/DM algorithm.

We seek the optimal solution through the evolution of a particle in the membrane structure and the exchange of an adjacent membrane structure among the optimal particles. Experimental results indicate that, in terms of time and effect, MOEA/DM performs better than MOEA/D.

The rest of this paper is organized as follows. Section 2 describes the community detection and concept of multi-objective optimization. Section 3 elucidates the proposed MOEA/DM. Section 4 presents the experimental studies. Section 5 concludes the study.

Clustering problem related background

Network Community detection based on the graph. A network is usually expressed as a graph structure, $G = G(N, V)$, where N represents nodes and V represents the relationships between the network's nodes. For a graph, $G = G(N, V)$ can also be expressed as an adjacent matrix, A. For every element a_{ij} of A,

$$\begin{cases} a_{ij} = w_{ij} & \exists \, L(i, j) \\ a_{ij} = 0 & ! \, \exists \, L(i, j) \end{cases} \tag{1}$$

$L(i, j)$ denotes that node i and node j are connected; w_{ij} represents the weight value of the two nodes.

The purpose of network community detection is to determine the characteristic similarities of the nodes in the network and then classify them. If a network is undirected and unweighted when an edge is connected between two nodes, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. The degree of node i is defined as $k_i = \sum_{j=1}^{n} A_{ij}$. However, the degree of node i can also be expressed as $k_i = k_{ij}^{in} + k_{ij}^{out}$, and expressed as follows:

$$\begin{cases} k_i^{in}(S_i) = \sum_{j \in S_i} A_{ij} \\ k_i^{out}(S_i) = \sum_{j \notin S_i} A_{ij} \end{cases} \tag{2}$$

For network we divide into m communities, $S = \{S_1, S_2, \ldots, S_m\}$, if $\forall i \in S_i, \sum_{i \in S_i} k_i^{in} > \sum_{i \in S_i} k_i^{out}$, this community is called strong sense, and community in a weak sense if $\forall\ i \in S_i, \sum_{i \in S_i} k_i^{out} > \sum_{i \in S_i} k_i^{in}$. The above statement shows that, in a strong community, each node has more connections within a community than with other communities in the network. In a weak community, the sum of the degrees within the rest of the community, is greater than the sum of the degrees within the community.

Multi-objective Optimization(MOP). A multi-objective optimization problem is stated as follows:

$$\begin{cases} minimize\ F(X) = (f_1(X), \ldots, f_m(X))^T \\ s \cdot t \quad X \in \Omega \end{cases} \tag{3}$$

where Ω is the variable space. $F : \Omega \to R^m$ contains m objective functions where R^m is defined as the objective space. Unlike a single objective optimization problem producing one optimal solution, there are probably many, even infinite, solutions for problem (3). These feasible solutions are called Pareto Optimality. Let $u, v \in R^m$. u is said to dominate v if for any $i \in \{1, 2, \ldots, m\}$, $u_i \geq v_i$ and there exists at least one $j \in \{1, 2, \ldots, m\}$ that $u_j > v_j$. If there is no point $X \in \Omega$ such that X dominates X^*, then X^* is a Pareto optimal solution. All the non-dominated X^* set is called Pareto Front. However, it is time-comsuming and even impossible to find the entire Pareto Front. Therefore, most algorithms aim to find out an even-distributed part of Pareto Front to represent the whole one.

Under certain conditions, a multi-objective optimization problem(MOP) can be decomposed into several single objective optimization problems(SOPs). There are two types of algorithms to decompose an MOP into a group of SOPs. The first type is weight aggregation based decomposition approaches, a set of weight vectors are used to convert an MOP into a number of SOPs using a scalarization method. the weighted Tchebycheff approach[27] and the PBI approach[21] are most widely used. The second type decomposes the objective space into a group of subspaces using a set of weight/reference vectors, which are most widely used in recent years.

The Tchebyshev method is classic and is expressed as:

$$min\ g^{te}(X|\lambda, Z^*) = \max_{1 \leq i \leq m} \lambda |f_i(X) - Z_i^*| \tag{4}$$

where $X \in \Omega$, z^* is the reference point. For $i \in [1, m]$, $Z_i^* = min\ f_i(x)$. $Z_i^* = min\{f_i(x)|x \in \Omega\}$, $i = \{1, 2, \ldots, m\}$. For each non-dominated solution x^* of (3), there exists a weight vector lamda so that x^* is the optimal solution of (4). We cannot conclude whether or not (1) the pareto front is concave and (2) the two objectives we use in this paper are discontinuous. When the pareto front is non-concave, the weighted sum approach does not work well, and that is why we choose the Tchebycheff approach.

MOEA/DM for Community Detection

Introduction of MOEA/DM algorithm. In 2004, MOEA/D was proposed. However, we discovered there is a lack of diversity in the pareto front. We assume the reason for this is there may be a number of sub problems corresponding to the same non-dominated solutions. Therefore, we propose a MOEA/DM algorithm to reduce the number of sub problems and improve the probability that the solution is not the same for each sub-problem and corresponding optimal.

Objective function. For an unsigned network, the degree of the node reveals the closeness between the nodes. Modularity density (D)[28], widely used in a variety of community detection algorithms, is one of the most basic measurement standards that uses the degree of the node. D is defined as:

$$D = \sum_{i=1}^{k} \frac{L(S_i, S_i) - L(S_i, \overline{S}_i)}{|S_i|} \tag{5}$$

In Eq. (5), each sum means the ratio between the difference of the internal and external degrees of the subgraph S_i and the size of the subgraph. In the above formula, we define $L(S_i, S_j) = \sum_{i \in S_i, j \in S_j} A_{ij}$ and $L(S_i, \overline{S}_j) = \sum_{i \in S_i, j \in \overline{S}_j} A_{ij}$. Give a partition $S = (S_1, S_2, \ldots, S_m)$ of the graph, where S_i is the vertex set of subgraph G_i ($i = 1, 2, \ldots, m$). However, MOEA/D divided D into two parts as two objectives one of which is NRA (negative ratio association) and the other one is ratio cut (RC).

$$min \begin{cases} NRA = -\sum_{i=1}^{k} \frac{L(S_i, S_i)}{|S_i|} \\ RC = \sum_{i=1}^{k} \frac{L(S_i, \overline{S}_i)}{|S_i|} \end{cases} \tag{6}$$

RC is used to measure the connection density between communities and NRA is used to measure connection density within communities. If these two goals are optimized simultaneously These two objectives can be determined to minimize the community more closely and the internal connection between communities sparse connection. Given a partition $S = (S_1, \ldots, S_m)$, S_i are the decision variables, and m is the scale (i.e., the number of decision variables) of the problem.

Encoding and decoding of discrete population position. Proposed in the graph structure is a genetics-based adjacency matrix notation[29], where each individual, g, of the population consists of N genes, each of which, takes allele values, j, in the range $1, 2, \ldots, N$. Genes and alleles represent nodes of the graph $G = (V, E)$
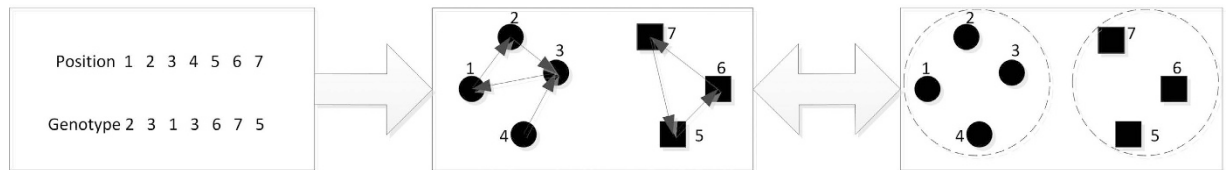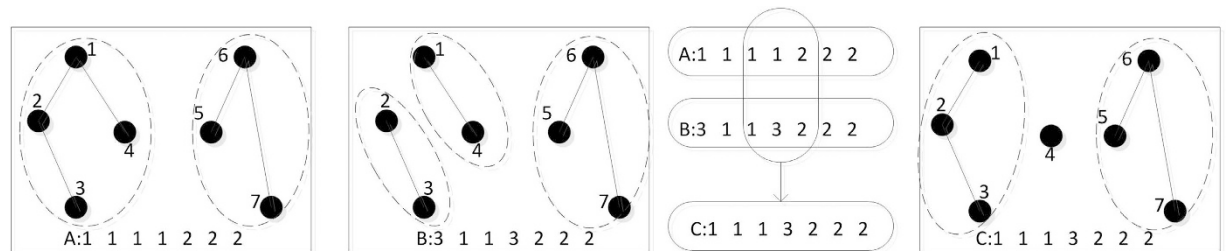
Figure 3. Encode of network.

Figure 4. Crossover of solution.

modeling a network. Thus, *a* value of *j*, assigned to the *i*th gene, is then interpreted as a link between the nodes *i* and *j*, and, in the resulting clustering solution, the nodes are in the same cluster. The decoding of this representation requires the identification of all connected components. All nodes belonging to the same connected component are then assigned to one cluster. A main advantage of this representation is that it is unnecessary to fix the number of clusters in advance, because the number of clusters is automatically determined in the decoding step. Figure 3 illustrates the locus-based adjacency scheme for a network of seven nodes.

Crossover. We choose the two-point crossover, in favor of uniform crossover, because the two-point crossover better maintains effective node connections in the network. Given two parents, A and B, we first randomly select two points i and j ($1 \leq i \leq j \leq N$), and then everything between the two points is swapped between the parents (i.e. $A_k \leftrightarrow B_k$, $\forall k \in \{k | i \leq k \leq j\}$. An example of the operation of two-point crossover on encoding is shown in Fig. 4.

Combination of Evolutionary Algorithms and Membrane Structure. The main idea of MOEA/DM is that object space is divided into a plurality of membrane structures and the solution of each membrane structure is initialized. Through population evolution within the membrane to screen out the best solution in each membrane and passed to the adjacent membrane structure. In each evolution of the membrane interior and also remove the worst performance of the solution. So in a sub-problem we choose the best solution is relatively more and to ensure that in each membrane is the best. Through a number of iterations, the solutions of each membrane structure is considered to be the best solutions of the sub-problem that corresponding membrane structure.

Through in comparison to the four known classification network and in two unknown classification network MOEA/D and MOPSO, MOEA/DM in the calculation of the cost of time is much faster than the other two algorithms and also in effect superior by. The algorithm flow can be expressed as Fig. 5.

Experimental Results

In this paper, we compare, mainly in time and $Q^{4,29}$ values, our proposed algorithm with one EA-based algorithm (MOEA/D) and one PSO-based algorithm (MOPSO). Experimental parameters are listed in Table 1. Compare the number of iterations of the three algorithms, those algorithms set for the 200 generation. Finally, we use the modularity proposed by Newman and Girvan. The modularity is defined as:

$$Q = \sum_{S=1}^{k} \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right]$$

(7)

MOEA/D algorithm. MOEA/D uses the Pareto dominance concept to allow the heuristic to handle problems. The algorithm provides a simple, but efficient, way of introducing decomposition approaches into multi-objective evolutionary computation. A decomposition approach, often developed in the community of mathematical programming, is readily incorporated into EAs in the MOEA/D framework to solve MOPS. Because MOEA/D is better than the EA-based algorithm, we chose MOEA/D for comparison with our proposed algorithm (MOEA/DM) to determine the performance of time cost and Q-value.

MOPSO algorithm. The MOPSO uses the same Pareto dominance concept to allow the heuristic to handle problems. However, the main difference between MOEA/D and MOPSO is the optimization strategy. Because MOPSO is better than the PSO-based algorithm, we chose MOPSO for comparison with our proposed algorithm (MOEA/DM) to determine the performance of time cost and Q-value.

Algorithm 1 Framework of the proposed MOEA/DM

Parameters: Max Generation:maxgen, **Number of cell**:cellNum, **Population**: pops, Mutation

Probability:pm, **inertia weight**: W, **Adjacent cell optimal solution**: best_solution.

Input:The adjacent matrix A of a network.

Output:The maximum value of Q, and the time spent Ctime.

Step 1:Initialization

Step 1.1:Generate a well_distributed weighted vectors:W $= \left\{ w_1, w_2, \cdots, w_{cellNum} \right\}^T$.

Step 1.2:Best solutions of adjacent cell:Best_solution $= \left\{ b_1, b_2, \cdots, b_{cellNum} \right\}$.

Step 2:Cycling

for $i = 1, 2, \cdots, pop.do$

Step 2.1:Cycling

for $j = 1, 2, ..., cellNum, do$

    Step:2.1.1:N particles are randomly generated in the cell.

    Step: 2.1.1:The optimal particle to add the adjacent cells in the current cell: b_j.

    Step 2.1.2: Intracellular particle hybridization.

    Step 2.1.3:Intracellular particle variation.

    Step 2.1.4:Update current cell optimal particle.

Step: 3:Output Q value and Ctime value.

Figure 5. Framework of the proposed MOEA/DM.

Algorithm	Parameters				
	Popsize	CellNum	Niche	PC	PM
MOEA/DM	120	40	13	1	0.06
MOPSO	100	NULL	40	1	0.06
MEA/D	100	NULL	40	1	0.06

Table 1. Experimental parameters.

For each test instance, both MOEA/D and MOPSO were run independently 100 times on the same computer (Inter(R) Celeron(R)M CPU 520 machine, 1.6 GHz, 512 MB memory). The operating system is Windows 8. In our experiments, the following performance indexes are used. There are many parameters that can be set flexibly, shown on Table 1, such as the parameter that are used to store the dominant solution, that we call it niche. The parameter niche is used to determine the neighborhood size and influence on the performance of our algorithm. In order to find the best value, we run the algorithm 200 times with different niche. After a lot of experiments, when niche equals 13 more suitable for this algorithm. Apart from the parameter niche, the population size popsize, the iteration number maxgen and the cell number CellNum are also affected the results of the experiment. We use the method of controlling variables, and ultimately determine their values. We set the popsize equal 120, the gmax equal 200 and the CellNum equal 40.

Experimental results on real-world networks. In this section, we demonstrate the MOEA/DM application effects on five real-world networks. Of these, the dolphin social[30], the American college football[31], the Zachary's karate club[32] and the political book network found from V. Krebs are known to be true. For the Santa Fe Institute SFI[33] and the netscience networks[34] the true data classification is unknown. The characteristics of the networks are given in Table 2. Table 3, we reflect on the performance of Q and running time on the value of the data that is known to the specific classification of MOEA/DM. Table 4, we reflect on the MOEA/DM performance of the Q value and time cost on the two unknown exact classification data.

Comparison of algorithms on the karate network. *Karate network.* The Karate network is a social network analysis in the field of classical data sets. In the early 1970s, the sociologist, Zachary, took two years to

Network	Parameters			
	Vertex	Edge	Real	Clusters
Karate	34	78	2	3
Dolphin	62	159	2	3
Football	115	613	12	12
Polbooks	105	882	3	8
SFI	118	200	unknown	13
Netscience	1589	2742	unknown	35

Table 2. Network Properties.

Network	MOEA/D		MOPSO		MOEA/DM	
	Q	Time(s)	Q	Time(s)	Q	Time(s)
Karate	0.4198	91.444	0.4198	216.901	0.4198	67.314
Dolphin	0.521	160.596	0.5265	548.290	0.5265	112.845
Football	0.6044	253.406	0.6033	1035.080	0.6046	174.158
Polbooks	0.5269	231.089	0.3600	990.917	0.5269	165.830

Table 3. A comparison of the data on the classification.

Network	MOPSO		MOEA/DM	
	Q	Time(s)	Q	Time(s)
SFI	0.7484	1683.24	0.7486	197.2360
Netscience	0.9503	26790.3	0.9518	578.2089

Table 4. Comparison of the data on the unknown.

observe the social relations among the 34 members of an American university karate club. Based on these internal club members as well as external exchanges, he constructed the social relations between members of the network consisting of 34 nodes. An edge between two nodes means that between the corresponding communities at least two members frequent exchanges of friends.

In Fig. 6(a), we show the true situation of the clustering karate network, In Fig. 6(b) we present the results of the clustering algorithm, MOEA/DM. In Fig. 6(b), MOEA/DM is divided into four categories: the top part is divided into two categories and the bottom part also divided into two parts. In Fig. 6(a), Point 10 (red) belongs to the real structure (upper part). According to our prediction, Point 10 (blue) should belong in the predicted structure (lower part) shown in Fig. 6(b). Other papers designate points (such as Point 10) as fussy nodes, i.e., it can be either classified to the first cluster or to the second one.

In the classification process, because points have just two edges to connect two different categories, points (such as Point 10) are divided into two parts. Although we used four categories (rather than two), we correctly divided the network.

Table 3 shows that, although the performance value index, Q, for our MOEA/DM is consistent with the values for MOEA/D and MOPSO, the time value for MOEA/DM is superior to the times for MOEA/D and MOPSO.

Comparison of algorithms on the dolphin network. *Dolphin network.* In New Zealand's life habits of 62 bottlenose dolphins, Lusseau[30] found the dolphin's interaction with a specific pattern, and constructed a social network containing 62 nodes. This dolphin network is naturally separated into two large groups: female and male. In Fig. 7(a), we show the true situation of the clustering dolphin network, In Fig. 7(b) we present the results of the clustering algorithm, MOEA/DM. In Fig. 7(b), MOEA/DM is divided into four categories: the top part is divided into two categories and the bottom part also divided into two parts.

Table 3 shows that, in terms of the Q value, MOEA/DM's performance is the same as that for MOPSO, and both (MOEA/DM and MOPSO), in terms of Q value, perform better than MOEA/D. However, as indicated in Table 3 and shown graphically in Fig. 8, in terms of running time, MOEA/DM has the advantage that its running time is substantially less than half that of MOPSO.

Comparison of algorithms on the football network. *Football network.* When Jantonio Turner[31] wanted to find more football highlights and discovered that no other all-football channel existed, he founded the football network in August, 1996. He was first mentored by Sheldon Altfeld, who had launched his own channel and who by then was giving seminars to entrepreneurs who wished to begin their own networks.

The network is divided into twelve categories as shown in Fig. 9(a). Figure 9(b) shows the classification results after using the MOEA/DM algorithm. A comparison of Fig. 9(a) with Fig. 9(b) shows that the football network has a more complex structure than the Dolphin and Karate networks. In the football network, nodes belong to the
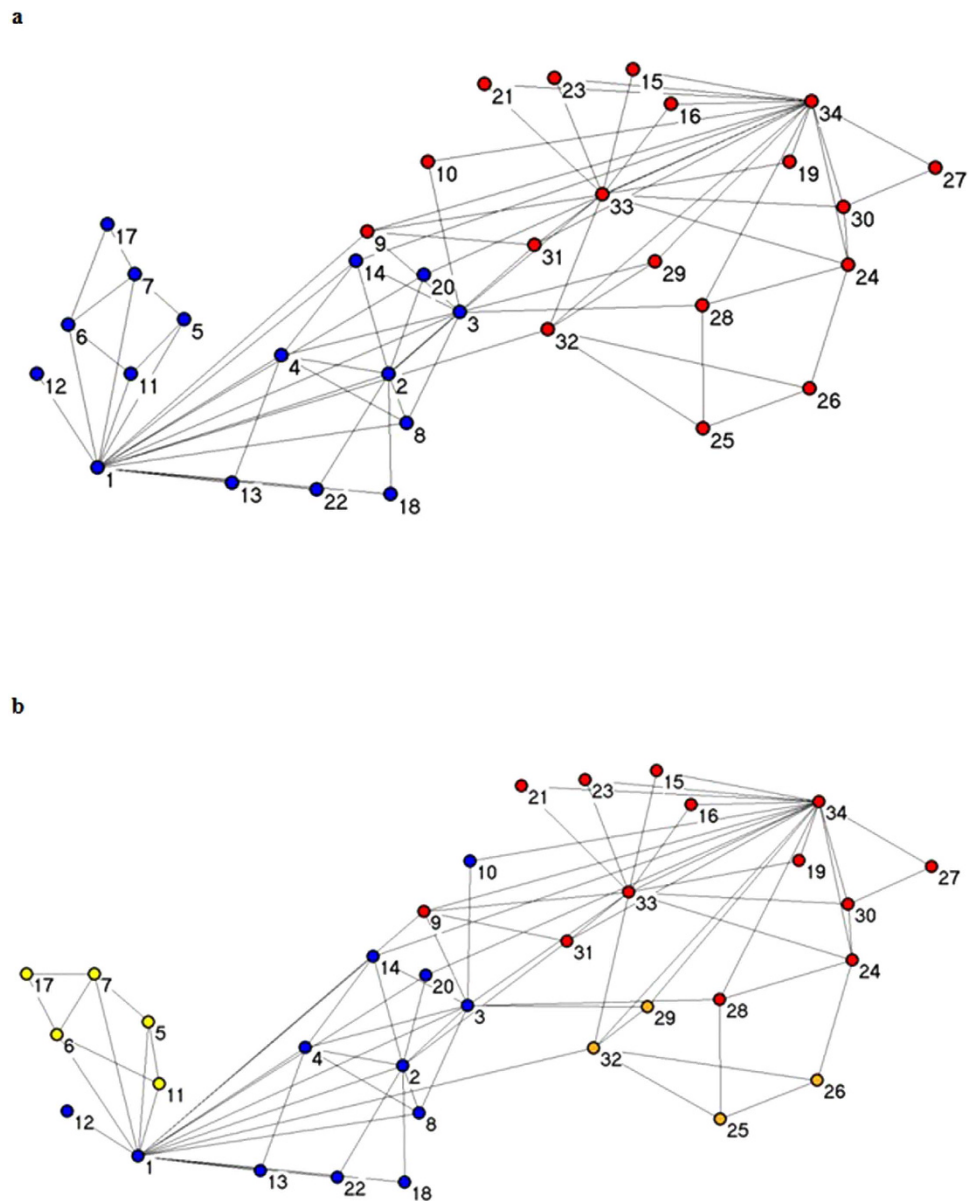
Figure 6. Clustering results on Karate club network by MOEA/DM. (**a**) The real structure of Karate network. (**b**) The prediction structure of Karate network detected by MOEA/DM.

same classare relatively decentralized. The real network structure, shown in Fig. 9(a), and our predicted network structure, shown in Fig. 9(b), have the same number of categories. From Fig. 9(b) we extracted the three categories on the right and placed them in Fig. 9(c). The three categories in Fig. 9(c) appear to classify the wrong point. The point that marked 58, 29 and the 43, 37, 91 be divided into the wrong position. An analysis of these points reveals that a characteristic they have in common is connecting to other classes is more prevalent than connecting to the edges of their own classes. The MOPSO algorithm divides the network into a like category, but more than 10 points are incorrectly placed.

Table 3 shows that, in terms of the Q value, MOEA/DM's performance is better than MOEA/D and MOPSO. As indicated in Table 3 and shown graphically in Fig. 8, in trems of running time, MOEA/DM has the advantage that its running time is substantially less than half that of MOPSO.

Comparison of algorithm on the polbooks network. *American political book network.* The American political book network, based on American political books, is a network of V. Krebs, which has been established on Amazon's online bookstore. Network edges represent that more readers bought two books simultaneously. This information is obtained from the purchase of books on the web page provided by the "purchase of the book's readers also buy books." At the same time, according to the point of view and evaluation of the readers of the Amazon books, Mark Newman divided the node types into three categories: "free," "conservative," and "centrist."
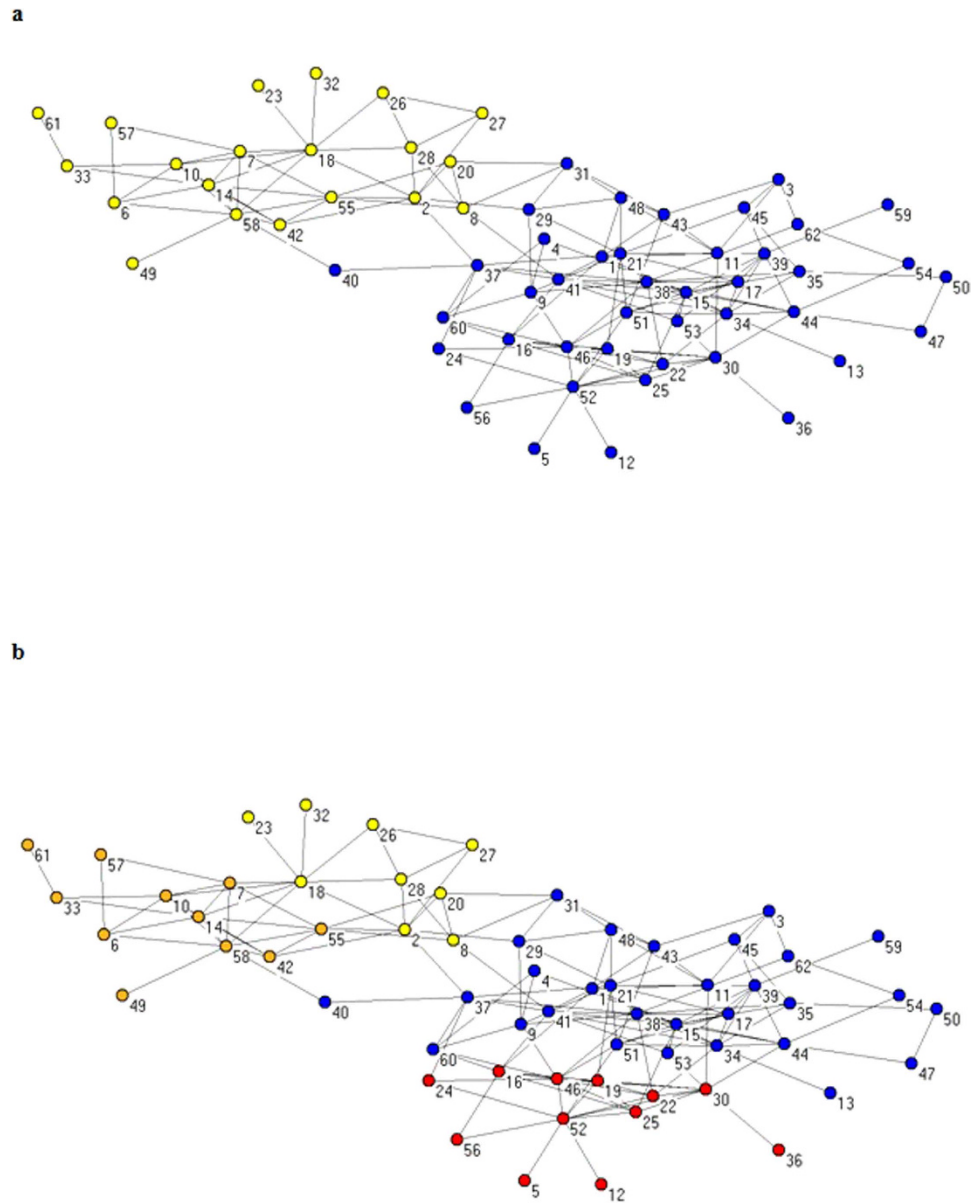
Figure 7. Clustering results on Dolphin club network by MOEA/DM. (**a**) The real structure of Dolphin network. (**b**) The prediction structure of Dolphin club network detected by MOEA/DM.

The network is divided into three categories as shown in Fig. 10(a). Figure 10(b) shows the classification results after using the MOEA/DM algorithm. A comparison of Fig. 10(a) with Fig. 10(b) shows that the political network has a more complex structure than the Football networks. In the political network, nodes belong to the same classare relatively decentralized. MOEA/DM divides the network into eight category and the part that the color mark red, have divides 9 points are incorrectly placed. The category that color marks blue, have divides four major sub category the color mark orange, green, pink and yellow.

Table 3 shows that, in terms of the Q value, MOEA/DM's performance is the same as that for MOEA/D, and both (MOEA/DM and MOEA/D), in terms of Q value, perform better than MOPSO. However, as indicated in Table 3 and shown graphically in Fig. 8, in terms of running time, MOEA/DM has the advantage that its running time is substantially less than MOEA/D.

Experimental results on unknown networks. The SFI[34] network represents 271 scientists in residence at the Santa Fe Institute, Santa Fe, NM, USA, during any part of calendar year 1999 or 2000, and their collaborators. An edge is drawn between a pair of scientists if they coauthored one or more articles during the same time period. The biggest component of the SFI graph consists of 118 vertices and we only do experiments on this part. Figure 11 show the result of the MOEA/DM. From the picture, the network be divided into twelve category and MOPSO divided into eight category. From the Table 4, the Q value result from MOEA/DM better than the result from MOPSO and the time cost less than MOPSO.
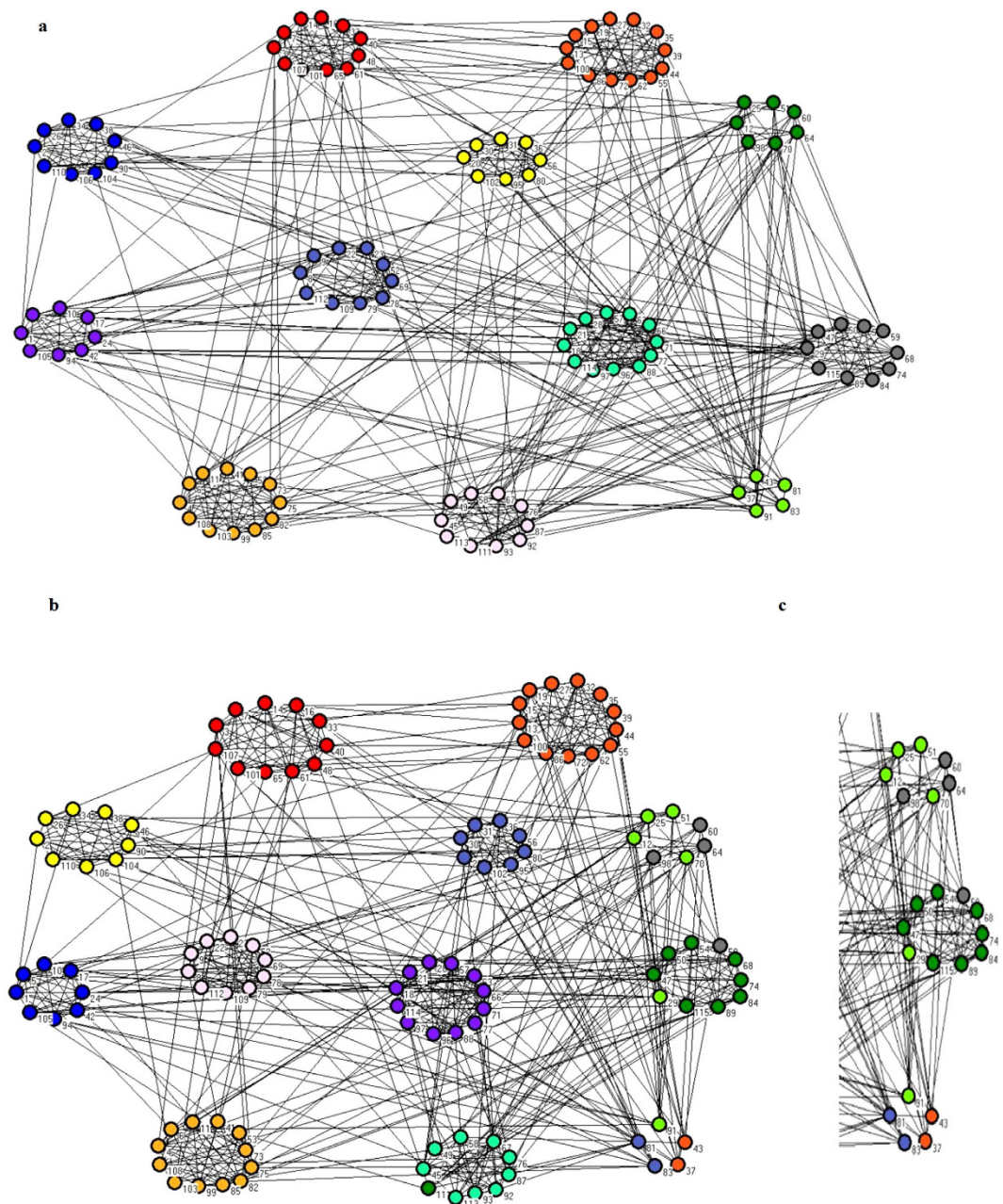
Figure 8. Comparison of three algorithms in time cost.

Figure 11 illustrates the results from the algorithm, MOEA/DM, and compare with the MOPSO. In Fig. 11, it very clearly that MOEA/DM splits the network into eight communities, with the same network the algorithm that MOPSO splits network into eight. From the top of Fig. 11 to the bottom, category at the top represents a group of scientists using agent-based models to study problems in economics and traffic flow we shows with the color blue. The second category represents a group of scientists working on mathematical models in ecology we shows with the color red. The third category which made up of four parts, the color in the picture is red, yellow, green, blue, represents a group of scientists working primarily in statistical physics. The two algorithm subdivide this group into four small ones. At the bottom of the figure is a group working primarily on the structure of RNA.

Concluding remarks

This study introduced an algorithm that combines membrane structure and an evolutionary algorithm, MOEA/DM. In the process of studying the MOEA/D algorithm, it is found that a non-dominated solution corresponds to multiple sub problems. MOEA/DM algorithm, mainly in the number of sub problems and the corresponding solution of each sub problem, improves the number of solutions of one sub problem by trying to reduce the
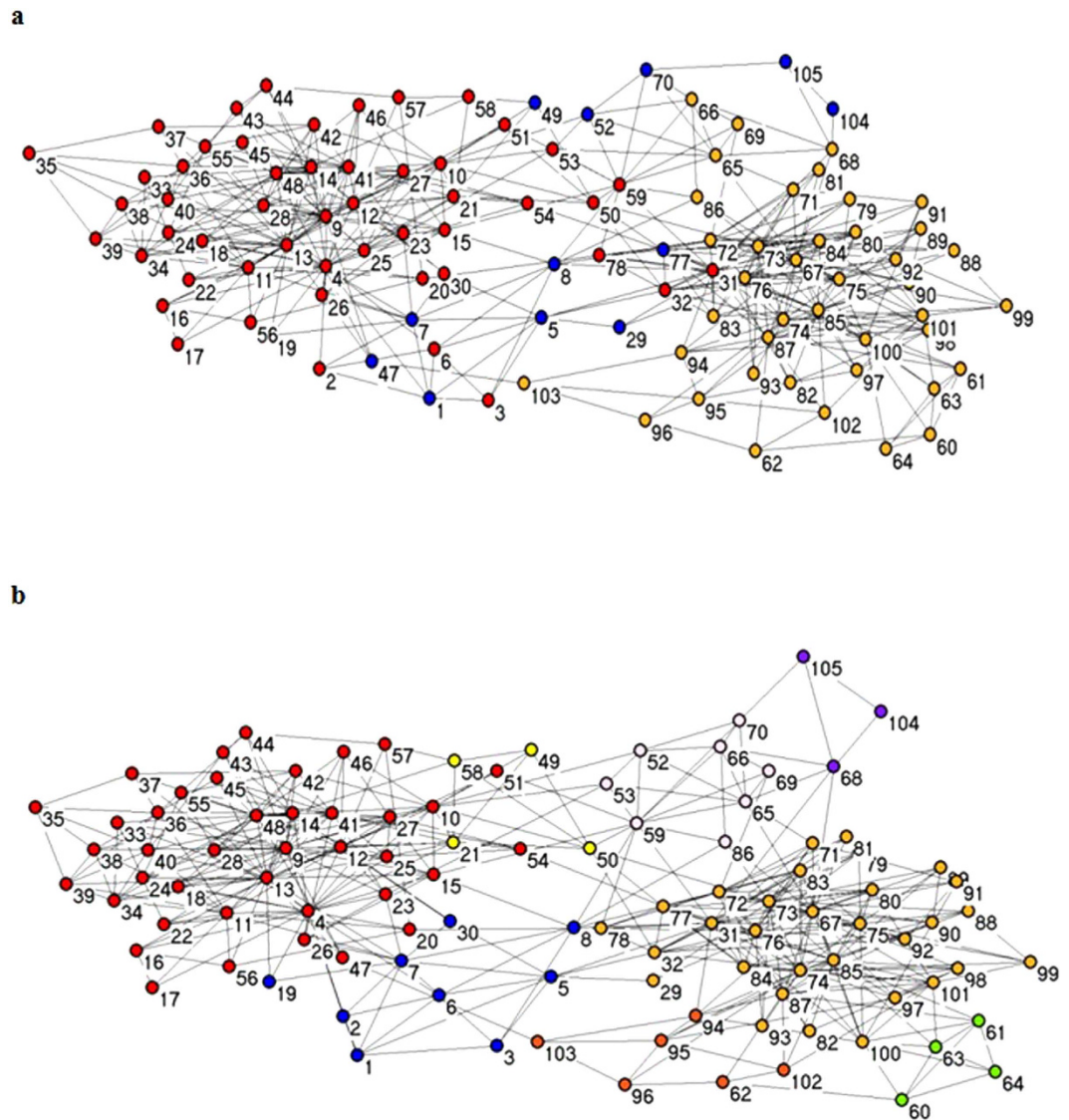
Figure 9. Clustering results on Football club network by MOEA/DM. (**a**) The real structure of Football network. (**b**) The prediction structure of Football network detected by MOEA/DM. (**c**) The apart of (**b**).

number of sub problems and the addition of film structure to try to ensure that each sub problem has a different number of solutions. Through experiments in the real network, it is found that this improvement has a certain effect. The following is a summary of the three improvements of MOEA/DM:

1. The diversity performance of the proposed algorithm is high because it places the target space on the average weight vector, and the membrane structure is divided into several parts.
2. The time efficiency of the proposed algorithm is higher than those of MOEA/D and MOPSO because the average algorithm to target by the proposed algorithm is divided into several parts: a few particles within the membrane of the evolutionary algorithm.
3. The effect of the proposed algorithm is better than those of MOEA/D and MOPSO, and spends much less time.

Figure 8 illustrates that MOEA/DM has a great advantage in running time. Nevertheless, the results of the experiment in a real network indicate that although MOEA/DM rapidly and accurately locates the real community, it inevitably produces errors in terms of the community number. If a corresponding estimate of the network category is obtained, then the effect is better. In addition to the category of problems, we also determined that certain points (known as the concepts of point classification) still present a high probability of error. Such results are usually generated in the connection between two communities with the same side. If the two goals can be optimized, better results may be obtained.
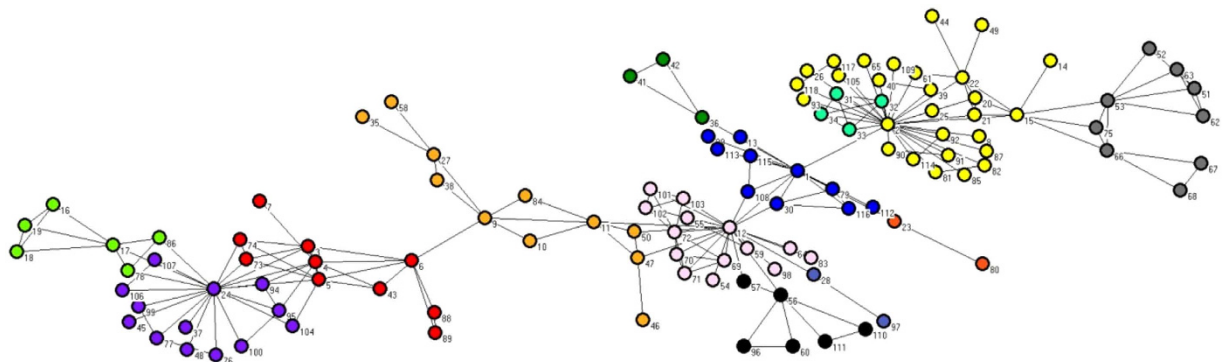
Figure 10. Clustering results on Polbooks club network by MOEA/DM. (**a**) The real structure of American political book network. (**b**) The structure of American political book network detected by MOEA/DM.
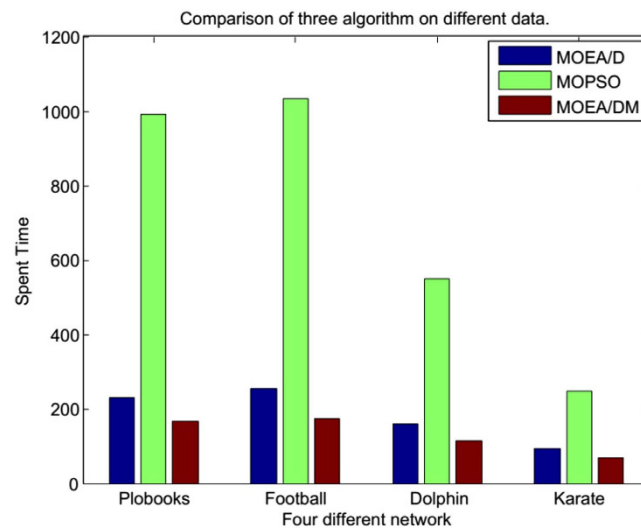

Figure 11. Prediction structure of SFI detected by MOEA/DM.

It is anticipated that the proposed algorithm for complex network clustering can be applied to the field of bioinformatics, such as Disease Gene network[35–37], DNA binding protein network identification[38,39], protein remote homology detection[40], etc. It is of interests to consider machine learning methods[41–45] for network clustering. Recently, spiking neural network models, see e.g. refs 46–48 particularly the ones with-self organizing[49,50] have been a hot topic in the field of machine learning, it is expected to obtain interesting result with this new powerful model.

References

1. Liu, Y., Zeng, X., He, Z. & Zou, Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Transactions on Computational Biology & Bioinformatics* 1–1 (2016).
2. Zou, Q. *et al.* Prediction of MicroRNA-Disease Associations Based on Social Network Analysis Methods. *Biomed Research International* **2015** (2014).
3. Zou, Q., Li, J., Wang, C. & Zeng, X. Approaches for Recognizing Disease Genes Based on Network. *Biomed Research International* **2014,** 416323–416323 (2014).
4. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 7821–7826 (2002).
5. Schuetz, P. & Caflisch, A. Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **77,** 357–361 (2008).
6. Wilson, E. B. The Distribution of the Chondriosomes to the Spermatozoa in Scorpions. *Proceedings of the National Academy of Sciences* **2,** 321–324 (1916).
7. Yuchun, R. Q. A comparison of rate control and rhythm control in patients with atrial fibrillation. *The Journal of Evidence-based Medicine* **347,** 1825–1833 (2004).
8. Fortunato, S. Community detection in graphs. *Physics* **486,** 75–174 (2010).
9. Cheng, R. & Jin, Y. A competitive swarm optimizer for large scale optimization. *IEEE Transactions on Cybernetics* **45,** 191–204 (2014).
10. Zhang, X., Tian, Y., Cheng, R. & Jin, Y. An Efficient Approach to Non-dominated Sorting for Evolutionary Multi-objective Optimization. *IEEE Transactions on Evolutionary Computation* **19,** 201–213 (2015).

11. Cheng, R. & Jin, Y. A social learning particle swarm optimization algorithm for scalable optimization. *Information Sciences* **291,** 43–60 (2015).
12. Zhang, X., Tian, Y. & Jin, Y. A Knee Point-Driven Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation* **19,** 761–776 (2015).
13. Barabasi, A. L. & Albert, R. Emergence of Scaling in Random Networks. *Science* **286,** 509–512 (1999).
14. Pizzuti, C. GA-Net: A Genetic Algorithm for Community Detection in Social Networks. Parallel Problem Solving From Nature - PPSN X, International Conference Dortmund, Germany, September 13–17, 2008. *Proceedings* 1081–1090 (2008).
15. Gong, M., Jiao, L., Du, H. & Bo, L. Multiobjective immune algorithm with nondominated neighbor-based selection. *Evolutionary Computation* **16,** 225–255 (2008).
16. Wu, Y., Luo, X., Zheng, F., Yang, S., Cai, S. & Ng, S. C. Adaptive Linear and Normalized Combination of Radial Basis Function Networks for Function Approximation and Regression. *Mathematical Problems in Engineering* **2014,** 1–14 (2014).
17. Wu, Y. *et al.* Quantification of knee vibroarthrographic signal irregularity associated with patellofemoral joint cartilage pathology based on entropy and envelope amplitude measures. *Computer Methods & Programs in Biomedicine* **130,** 1–12 (2016).
18. Zhang, X., Tian, Y. & Jin, Y. Approximate non-dominated sorting for evolutionary many-objective optimization. *Information Sciences* (2016).
19. Cheng, R., Jin, Y., Olhofer, M. & Sendhoff, B. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation* (2016).
20. Cheng, R., Jin, Y., Narukawa, K. & Sendhoff, B. A Multiobjective Evolutionary Algorithm Using Gaussian Process-Based Inverse Modeling. *IEEE Transactions on Evolutionary Computation* **19,** 838–856 (2015).
21. Zhang, Q. & Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation* **11,** 712–731 (2007).
22. Coello, C. A. C. & Lechuga, M. S. MOPSO: a proposal for multiple objective particle swarm optimization. *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*, 1051–1056 (2002).
23. Zhang, X., Liu, Y., Luo, B. & Pan, L. Computational power of tissue P systems for generating control languages. *Information Sciences* **278,** 285–297 (2014).
24. Zhang, X., Zeng, X., Luo, B. & Pan, L. On Some Classes of Sequential Spiking Neural P Systems. *Neural Computation* **26,** 974–997 (1989).
25. Zhang, X., Pan, L. & Paun, A. On the Universality of Axon P Systems. *IEEE Transactions on Neural Networks & Learning Systems* **26,** 2816–2829 (2015).
26. Zhang, X., Wang, B. & Pan, L. Spiking neural p systems with a generalized use of rules. *Neural Computation* **26,** 2925–2943 (2014).
27. Temlyakov, V. Chebushev Greedy Algorithm in convex optimization. *Mathematics* (2013).
28. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (1998).
29. Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **69,** 026113–026113 (2004).
30. Lusseau, D. *et al.* The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology & Sociobiology* **54,** 396–405 (2003).
31. Kim, J. H., Zong, W. G. & Kim, E. S. Parameter estimation of the nonlinear muskingum model using harmony search 1. *Jawra Journal of the American Water Resources Association* **37,** 1131–1138 (2007).
32. Lancichinetti, A. & Fortunato, S. Community detection algorithms: a comparative analysis. *Physical Review E Statistical Nonlinear & Soft Matter Physics* **80,** 2142–2152 (2009).
33. Zhang, Z. *et al.* High throughput sequencing recovers the noval gene Megf8 as the ENU induced mutation causing single ventricle cardiac defect associated with randomized left-right patterning (Massively parallel sequencing identifies the gene Megf8 with ENU-induced mutation causing heterotaxy). *Proceedings of the National Academy of Science* 3219–3224 (2009).
34. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large networks. *Phys Rev E* **70(6),** 066111 (2005).
35. Zeng, X., Zhang, X. & Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Briefings in Bioinformatics* **17** (2015).
36. Quan, Z., Li, J., Li, S., Zeng, X. & Wang, G. Similarity computation strategies in the microRNA-disease network: a survey. *Briefings in Functional Genomics* **15** (2015).
37. Zeng, X., Liao, Y., Liu, Y. & Zou, Q. Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Transactions on Computational Biology & Bioinformatics* 1–1 (2016).
38. Liu, B., Wang, S., Dong, Q. & Li, S. Identification of DNA-binding proteins by combining auto-cross covariance transformation and ensemble learning. *IEEE Transactions on Nanobioscience* 1–1 (2016).
39. Liu, B. *et al.* iDNA-Prot|dis: identifying DNA-binding proteins by incorporating amino acid distance-pairs and reduced alphabet profile into the general pseudo amino acid composition. *Plos One* **9,** e106691–e106691 (2014).
40. Liu, B., Chen, J. & Wang, X. Application of Learning to Rank to protein remote homology detection. *Bioinformatics* **31,** 3492–3498 (2015).
41. Gu, B., Sheng, V. S., Tay, K. Y., Romano, W. & Li, S. Incremental Support Vector Learning for Ordinal Regression. *IEEE Transactions on Neural Networks & Learning Systems* **26,** 1403–1416 (2014).
42. Zheng, Y., Byeungwoo, J., Xu, D., Wu, Q. M. J. & Zhang, H. Image segmentation by generalized hierarchical fuzzy C-means algorithm. *Journal of Intelligent & Fuzzy Systems* **28,** 4024–4028 (2015).
43. Wen, X., Shao, L., Xue, Y. & Fang, W. A rapid learning algorithm for vehicle classification. *Information Sciences* **295,** 395–406 (2015).
44. Ma, T. *et al.* Social Network and Tag Sources Based Augmenting Collaborative Recommender System. *Ieice Transactions on Information & Systems* **E98.D,** 902–910 (2015).
45. Gu, B. & Sheng, V. S. A Robust Regularization Path Algorithm for ν-Support Vector Classification. *IEEE Transactions on Neural Networks & Learning Systems* 1–8 (2016).
46. Song, T., Xu, J. & Pan, L. On the Universality and Non-universality of Spiking Neural P Systems with Rules on Synapses. *IEEE Transactions on Nanobioscience* **14,** 1–1 (2015).
47. Song, T. & Pan, L. Spiking Neural P Systems with Request Rules. *Neurocomputing* **193,** 193–200 (2016).
48. Zeng, X., Xu, L., Liu, X. & Pan, L. On languages generated by spiking neural P systems with weights. *Information Sciences* **278,** 423–433 (2014).
49. Wang, X., Song, T., Gong, F. & Zheng, P. On the Computational Power of Spiking Neural P Systems with Self-Organization. *Scientific Reports* **6** (2016).
50. Zeng, X., Zhang, X., Song, T. & Pan, L. Spiking neural P systems with thresholds. *Neural Computation* **26,** 1340–1361 (2014).

Acknowledgements

Author Contributions

Y.J. put forward the idea of the paper. Xx.Z. and Xy.Z. conceived and designed the study. S.Z. performed the experiments and wrote the paper. N.D., Xx.Z. and Xy.Z. reviewed and edited the manuscript. All authors read and approved the final manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ju, Y. *et al.* Complex Network Clustering by a Multi-objective Evolutionary Algorithm Based on Decomposition and Membrane Structure. *Sci. Rep.* **6**, 33870; doi: 10.1038/srep33870 (2016).