Multi-objective Decisionmaking in the Detection of Comprehensive Community Structures

Chuan Shi*, Zhenyu Yan†, Xin Pan*, Yanan Cai* and Bin Wu*
* Beijing University of Posts and Telecommunications, Beijing, China 100876
shichuan@bupt.edu.cn.
†Research Department, Fair Isaac Corporation, San Rafael, CA, USA, 94903
yan_zhen_yu@hotmail.com.

Abstract—**Community detection in complex networks has attracted a lot of attentions in recent years. Compared with the traditional single-objective community detection approaches, the multi-objective approaches based on evolutionary computation can provide a decision maker with more flexible and promising solutions. How to make effective use of the optimal solution set returned by the multi-objective community detection approaches is an important yet unsolved issue. Through leveraging an existing multi-objective community detection algorithm, this paper proposes four model selection methods to aid the decision makers to select the preferable community structures. The experiments with three synthetic and real social networks illustrate that the proposed method can discover more authentic and comprehensive community structures than traditional single-objective approaches.**

Index Terms—**Complex network, community detection, multi-objective ptimization, evolutionary computation**

I. Introduction

Analysis of large complex networks, such as social network, World Wide Web, have drawn great interests in various research communities. This topic is important because those communities often play special roles in the network systems. Detecting the community structure in a complex network helps better understand the network system and thus has practical applications.

Many methods and algorithms have been developed for Community Detection (CD) [3], [4]. Most contemporary community detection algorithms choose a cost function, such as modularity Q [1] and "cut" [2] function, which measures the quality of community partitions first, and then optimize this function through searching the solution space. These algorithms can be regarded as single-objective methods. That is, a single-objective function is designed beforehand and the algorithm returns a single solution as results.

Although these single-objective approaches achieve great successes in both artificial and real networks, they have some fundamental drawbacks. For example, they often cause a fundamental discrepancy that different algorithms may produce distinct solutions for the same network. Moreover, these approaches have the resolution limit problem [6], that is, modularity optimization fails to find small communities in large networks.

In order to alleviate disadvantages in single-objective community detection algorithms, a natural approach may be to consider community detection as a multi-objective optimization problem. Moreover, some evolutionary multi-objective community detection algorithms have been developed recently [21] [22]. These algorithms simultaneously optimize multi-objectives and return a set of optimal solutions. These multi-objective methods have preliminarily shown their advantages in detecting more accurate community structures. However, it is still an unsolved issue to make best use of these optimal solutions. These solutions are generated by the different trade-offs of the objectives, and they reveal different community structures from different perspectives. It is promising to detect more accurate and comprehensive structures through exploiting the tradeoffs among these optimal solutions.

This paper adapts a multi-objective community detection algorithm proposed by Shi et al. in [21]. This method generates a set of optimal solutions. Furthermore, three model selection criteria and a Possibility Matrix method are proposed to divide the set of optimal results into four parts according to their qualities, which not only assists the Decision Makers(DMers) to select the preferable ones from them, but also helps reveal more complex and real community structures. Experiments on synthetic and real networks show that the multi-objective method with the proposed model selection criteria can discover more authentic and comprehensive (e.g., hierarchical and overlapping) community structures than traditional single objective approaches.

This paper is arranged as follows. Section 2 introduces the related work. Section 3 describes the multi-objective community detection algorithm and the model selection methods. The experiments on artificial and real networks are done to validate the effectiveness and efficiency of the algorithm in Section 4. Section 5 concludes the paper.

978-1-4244-7835-4/11/$26.00 ©2011 IEEE

II. Related Work

Many different algorithms have been designed to analyze the community structure in complex networks. The algorithms use methods and principles of physics, artificial intelligence, graph theory and even electrical circuits [4]. One of the most known algorithms proposed so far is the Girvan-Newman (GN) algorithm that introduces a divisive method by iteratively cutting the edge with the greatest betweenness value [1]. Some improved algorithms have been proposed [11], [12]. These algorithms are based on a foundational measure criterion of community, modularity Q, proposed by Newman [1]. The larger the value of Q the more accurate a partition into communities is. As a consequence, the community detection becomes a modularity optimization problem. Because the search for the optimal (largest) modularity value is an *NP*-complete problem [13], many heuristic search algorithms have been applied to solve the optimization problem, such as extremal optimization [14], simulated annealing [3] and genetic algorithm [15].

Some other criteria are also used as the optimization objective. The Hamiltonian-based method introduced by Reichardt and Bornholdt (RB) [7] is based on considering the community indices of nodes as spins in a *q*-state Potts model. Recently, Arenas, Fernandez and Gomez (AFG) [8] proposed a multiple resolution procedure that allows the modularity optimization to go deep into the structure. These methods vary the thresholds by using a tuning parameter in their criteria and investigate the community structure at various resolutions. The modularity Q can be regarded as a special case of these two criteria. In addition, Fosvall and Bergstrom [5] proposed an information-theoretic foundation for the concept of modularity in networks, in which the network is composed of modules by finding an optimal compression of its topology. Although these criteria could effectively assess the quality of the community, the recent research show that the optimization based on single criterion has a fundamental disadvantage [6], [9]. That is, the optimization based on single criterion may fail to identify modules smaller than a scale which depends on the total size of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined. The genetic algorithm (GA), as an effective optimization technique, has also been used for community detection. In order to optimize the modularity Q, the GAs in ref. [15] and [18] use the cluster centers and the locus-based adjacency as the encoding scheme, respectively. Pizzuti proposes another GA to optimize the ”community score” criteria [19], [20]. These algorithms have the advantage that the number of communities can be automatically determined during the evolutionary process. However, these algorithms also have the resolution limit, because a single objective is applied.

More recently, some researchers regard the CD as a Multi-objective Optimization Problem (MOP) and solve the MOP with the Multi-Objective Evolutionary Algorithm (MOEA) [21], [22]. Pizzuti [22] proposed MOGA-Net to optimize the community score and community fitness. Shi et al. [21] proposed a MOEA to optimize two components of modularity Q. The two multi-objective methods show their benefits in detecting more accurate community structures. However, they did not further explore the benefits of multi-solutions returned by multi-objective methods.

III. Multi-objective community detection algorithm and its Model Selection

This section proposes a multi-objective evolutionary algorithm for community detection. The approach consists of two phases. The first community detection phase adapts a multi-objective evolutionary algorithm to discover communities, and returns a set of optimal solutions. The second model selection phase proposes three community selection criteria to assist DMers' decision making.

A. Community Detection Phase

In the community detection phase, we adapt a multi-objective community detection algorithm, MOCD [21], to generate a set of Pareto optimal solutions. We make a brief description here. They are described in detail in [21].

Algorithm framework The Pareto Envelope-based Selection Algorithm version 2 (PESA-II) [17], is used to form the basis of MOCDs community discovery phase. PESA-II follows the standard principles of an EA with the difference that two populations of solutions are maintained: an internal population (IP) of fixed size, and an external population (EP). The IP explores new solutions and achieves these by the standard EA process of reproduction and variation. The EP is to exploit good solutions by maintaining a large and diverse set of the non-dominated solutions discovered during search. Selection occurs at the interface between the two populations, primarily in the update of EP. The detailed implementation can be seen in ref. [17].

Objective function The objective functions quantify the optimality of a solution, so we should select optimization objectives that reflect the fundamentally different aspect of a good community partition. One of the most important objective functions is the modularity [1] which is defined as follows:

$$Q(C) = \sum_{c \in C} [\frac{|E(c)|}{m} - (\frac{\sum_{v \in c} deg(v)}{2m})^2] \tag{1}$$

where the sum is over the modules of the partition, $|E(c)|$ is the number of links inside module c, m is the total number of links in the network, C is a partition result, and $deg(v)$ is the degree of the node v in module c. We transform the Q function into two parts, $intra(C)$ and $inter(C)$, which are used as two objectives for the MOEA. These two conflict objectives reflect different aspects of a community.

$$\begin{aligned} & intra(C) = 1 - \sum_{c \in C} \frac{|E(c)|}{m} \\ & inter(C) = \sum_{c \in C} \left(\frac{\sum_{v \in c} deg(v)}{2m}\right)^2 \\ & Q(C) = 1 - intra(C) - inter(C) \end{aligned} \tag{2}$$

Since the $inter$ objective function increases with larger number of communities while the $intra$ objective function decreases, MOEA can keep the community number dynamic and avoid convergence to trivial solutions(the detail analysis can be seen in ref.[16]). Though more objective functions can be used, our experiments indicate that additional objective functions do not necessarily lead to better solutions. We will explore more possible objective functions in future work.

Genetic representation and its operations The locus-based adjacency encoding scheme is used as the genetic representation. In this graph-based representation, each genotype g consists of n genes $g_1, g_2, \ldots, g_n$ and each g_i can take one of the adjacent nodes of node i. Thus, a value of j assigned to the ith gene, is then interpreted as a link between node i and j. In the resulting solution, they will be in the same community. We choose the locus-based adjacency encoding scheme due to its many advantages. For one part, there is no need to fix the number of communities in advance. In contrast, many other methods need prior knowledge to set the number of communities. For another, compared with the former genetic algorithms [15], the search space is reduced to $O(d^n)$, where d is the degree of nodes.

The crossover operation in MOCD is done by intersecting two chromosomes selected randomly from the population. For simplicity, the two chromosomes are called source and destination, respectively. Firstly, a gene is selected randomly from the source chromosome, and then we iteratively search the gene values that the gene link to and transfer these values in source chromosome to the corresponding genes in the destination chromosome. The exchange of gene segments is bidirectional. The crossover operator is prone to replicate the good structures generated by evolution to the new individual. Moreover, it is able to effectively generate the individual with different structures. The operator's computational complexity is $O(l)$ (where l is the length of the gene segment, namely the size of the community selected). l is usually smaller than n. In the mutation operation, we randomly select some genes and assign them with other randomly selected adjacent nodes.

B. Model Selection Phase

MOCD does not return a single solution, but a set of Pareto optimal solutions. These community partitions correspond to different tradeoffs between the two objectives and also consist of communities of different sizes. Domain expertise can be leveraged to make the final decision through analyzing the alternative solutions. This is crucial for a problem with unknown structure, like CD. In addition, the DMer may desire that the set of candidate solutions can be further narrowed or some representative ones can be recommended. In this section, we therefore introduce some methods for assessing the quality of the Pareto optimal partition solutions. These methods are able to further identify some promising partitions from the optimal solutions.

Formally, let $CSet$ be the set of community partitions (i.e., the optimal solutions set returned by MOCD), C be a partition in $CSet$, and there are k communities in the partition $C: C = c_1 \cup c_2 \cup ... \cup c_k$. A partition result is also called a clustering model M.

Maximum Q criterion. The criterion selects the model with maximum modularity Q. Because of the relationship of Q and two objective functions (see Equation 2), it is easy to select the model with maximum Q, and the corresponding model is called M_Q.

$$S_{Max-Q} = \underset{C \in SF}{\operatorname{argmax}} \{1 - intra(C) - inter(C)\} \tag{3}$$

Strong community criterion. According to the strong community definition given by Radicchi et al. [11], each node i in each community c is validated whether to satisfy the strong definition. If the ratio of communities satisfying the strong definition is larger than the predefining strong community threshold λ_{str}, the corresponding partition result is called strong community, and the set comprising the strong communities is called StrMSet. (k_i^{in}(c) is the number of edges connecting node i to other nodes belonging to c. k_i^{out}(c) is the number of connections toward nodes in the rest of the network.)

$$\begin{aligned} & StrCSet = \{c | k_i^{in}(c) > k_i^{out}(c) \quad \forall i \in c\} \\ & StrRatio(C) = \frac{|StrCSet|}{|C|} \\ & StrMSet = \{C | StrRatio(C) > \lambda_{str}\} \end{aligned} \tag{4}$$

Weak community criterion. Similarly, according to their weak community definition [10], for each partition result, each community could be verified whether to satisfy the weak definition. If the ratio of communities satisfying the weak definition is larger than the predefining weak community threshold λ_{weak}, the corresponding partition result is called weak community, and the set comprising the weak communities is called WeakMSet.

$$WeakCSet = \{c | \sum_{i \in c} k_i^{in}(c) > \sum_{i \in c} k_i^{out}(c)\}$$
$$WeakRatio(C) = \frac{|WeakCSet|}{|C|} \quad (5)$$
$$WeakMSet = \{C | WeakRatio(C) > \lambda_{weak}\}$$

In the definitions, two parameters λ_{str} and λ_{weak} need to be settled in the range from 0 to 1 beforehand, and they control the size of StrMSet and WeakMSet. If the networks have obvious community structures, these two parameters are settled with large values, or otherwise with small values. These three criteria reflect the quality of solutions to some extent. Generally speaking, the community partition M_Q is commonly used measure, to which the DMer should pay more attention. Meanwhile, the solution in StrMSet and WeakMSet also provide much valuable structure information to the DMer.

In order to illustrate the statistical characteristics of multi-solutions visually, a probability matrix is introduced to describe the probability of a pair of nodes in same community.

Probability Matrix.The rows and columns of the matrix correspond to the indices of nodes. For a partition solution, if two nodes are in the same community, the corresponding matrix value is 1, or else it is 0. For multi-solutions, their results are accumulated as a Possibility Matrix. The matrix can be converted to a gray graph in which the higher probability corresponds to the darker gray.

IV. Experiments

We validate the effectiveness of MOCD through a synthetic hierarchical network, a synthetic overlapping network and a real social networks. The experiments are carried out on a 3GHz and 1G RAM computer running Windows XP.

A. Hierarchical Network

The hierarchical network is a K40-4 network consisting of a ring of cliques, connected through single link. The network has 40 cliques, and each clique is a complete graph with 4 nodes and 6 links. In the network, it is clear that there are 40 unit communities and the connected cliques can also be considered as a community. The network has been used in [6].

We run MOCD with the following parameters: the population size is 100, the running generation is 100, the crossover ratio is 0.6, and the mutation ratio is 0.4, λ_{str} and λ_{weak} both are 1. We also run two popular single-objective approaches on the network: the betweenness-based heuristic algorithm GN [1] and the GA-based modularity optimization algorithm GACD [18]. Note that GACD has the same parameters with MOCD. In this experiment, the running times of MOCD, GN, and GACD are 26, 41, and 21 seconds, respectively.

GN obtains a solution with 16 communities, and GACD reaches the maximal Q value 0.881 with 15 communities. In both solutions, some connected cliques are combined. According to the construction process, these two solutions can both be regarded as the correct partitions. However, they both fail to reveal the hierarchical characteristic of the network. MOCD obtains 100 non-dominated solutions which are illustrated in Fig.1(a). Please note that the inter and intra values are normalized (it is same in the following section). There are 78 correct partitions in M_Q and StrCSet with the number of communities from 26 to 40. There are two special models in these solutions. As illustrated in Fig.1(b), the model M_Q reveals the 26 communities with the highest granularity. Another special strong community solution (labeled II), shown in Fig.1(c), reveals all 40 cliques with the lowest granularity. Most solutions lie between these two solutions. The Possibility Matrix of all solutions are illustrated in Fig.1(d). We can clearly find the hierarchical structure in which some large communities may contain some connected cliques. Compared to one solution with the higher granularity returned by GN and GACD, MOCD can find the communities with different scales in one run, which reveals more structural information.

Using the experimental data, we analyze the relationship of the objective values and the number of communities as shown in Fig.2(a). It is obvious that with the increase of the number of communities, the inter values increase, whereas the intra values decrease. It validates that the two objective functions are conflicting and complementary and the modularity Q is a trade-off between these two objectives. As for the Q value, it seems to decrease with the increase of the number of communities. In order to observe their relationship more clearly, the relationship of the number of communities and Q values of solutions in StrMSet is shown in Fig.2(b). It is clear that with the increase of the number of communities the Q value trends to become small. As the experiments illustrated, the single-objective approaches (e.g., GN and GACD) could only reveal the communities with large sizes. In fact, all the community partitions with small sizes discovered by MOCD are also correct. The experiment further confirms the resolution limit in the approaches with single objective [6]: methods based on optimizing the modularity measure or other single criterion may fail to identify modules smaller than some thresholds. Comparing to those single-objective approaches, MOCD can discover the hierarchical network with different scales (i.e., both small and large sizes).

B. Overlapping Network

The second experiment is an example of overlapping network. The network consists of two large communi-

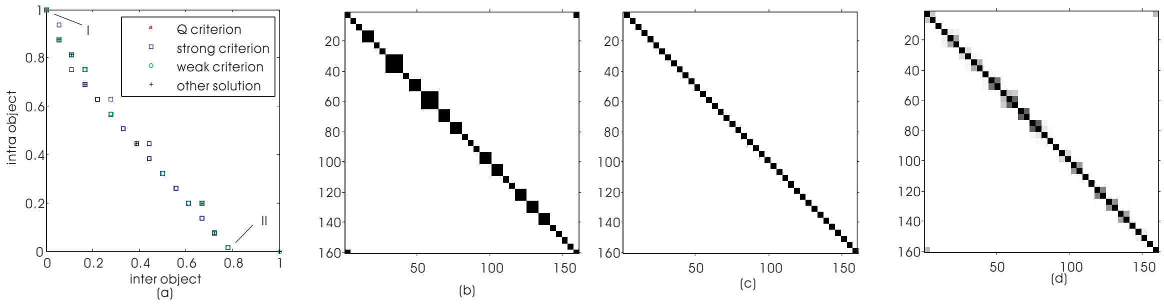

Fig. 1. Multiple resolution of modular structure in K40-4 network. (a) shows the curve of non-dominated solutions. (b) shows the Possibility Matrix of the solution labeled I with a gray graph. (c) shows the Possibility Matrix of the solution labeled II. (d) shows the Possibility Matrix of all solutions.

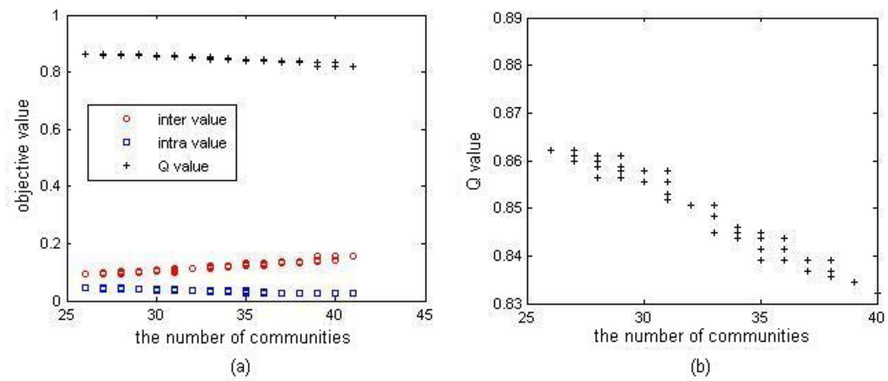

Fig. 2. Show the relationship of the number of communities and the objective values. (a) shows the relationship on the optimal solution set. (b) shows the relationship on the strong community set.

ties A and B, each containing 128 nodes, which have on average 12 internal links per node. Within A and B, a subgroup of 32 nodes exists, which we denote by a and b, respectively. Every node within this subgroup has six of its 12 intra community links with the 31 other members of this subgroup. The two subgroups a and b have on average three links per node with each other. Additionally, every node has one links with randomly chosen nodes from the network. It is clear that the network has two large communities (i.e., A and B) and one overlapping community a&b between A and B. The similar network has been used in [7].

MOCD settles the following parameters: the population size is 200, the running generation is 500, the crossover ratio is 0.6, and the mutation ratio is 0.4, λ_{str} is 0.3 and λ_{weak} is 0.5. GN and GACD are also run on this network, and GACD is equipped with the same parameters in MOCD. The running time of MOCD, GN and GACD are 214, 312, and 198 seconds, respectively.

GN and GACD both reveal the the large communities A and B accurately. However, they are not able to discover the overlapping structure. MOCD obtains 200 non-dominated solutions which are illustrated in Fig.3(a). All the solutions are divided into four types: 1 solution for M_Q, 9 solutions in StrCSet, 22 solutions in WeakCSet, and 165 other solutions. We also find two special clustering models in the figure. The model M_Q (labeled I in Fig.3(a)) reveals the same community structure as that of GN and GACD (i.e., two large communities A and B as illustrated in Fig.3(b)). Another special partition is a strong community solution (labeled II) as shown in Fig.3(c). The partition consists of three communities: two large communities, an overlapping community that is constituted by the nodes in a and b. The result shows that MOCD not only finds the obvious large community structure, but also reveals the implicit overlapping community in one run.

The overlapping community can actually be easily identified through an aggregation of all the solutions obtained from MOCD. Fig.3(d) shows the Possibility Matrix of all solutions of MOCD. We can see that an overlapping community which spans from node 98 to node 160 lies between the two large communities. A single solution obtained by any single-objective approach, such as GN or GACD can hardly discovers the overlapping structures. Whereas, the aggregation of all the optimal solutions obtained by MOCD can reveal it statistically. In ref. [7], Reichardt and Bornholdt also

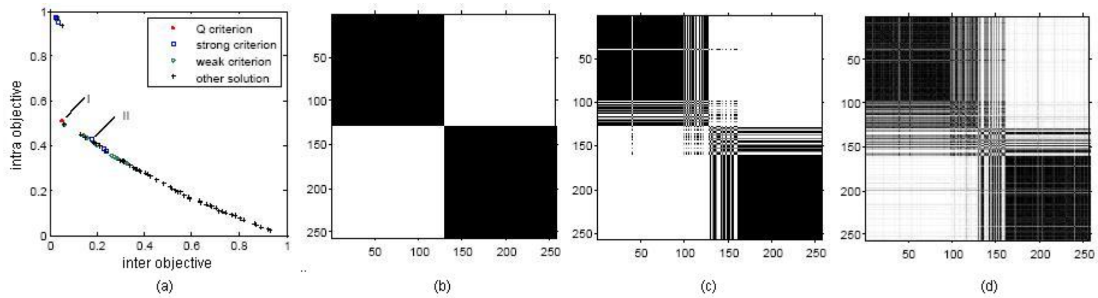

Fig. 3. Multiple resolutions of modular structure in the overlapping network. (a) shows the curve of non-dominated solutions. (b) shows the Possibility Matrix of the solution labeled with I with a gray graph. (c) shows the Possibility Matrix of the solution labeled with II. (d) shows the accumulated Possibility Matrix of all solutions.

found the two partitions in Fig.3(b) and (c) at $\gamma = 0.5$ and $\gamma = 1$, respectively. However, in order to discover the correct partition, many runs should be done to find the proper γ. Compared with their method, MOCD obtains many partitions including the correct partitions in one run and the aggregation of all the solutions is able to statistically reveal the hidden but informative structure.

C. Real Network

We now turn to a real world example to see whether these structural properties can indeed be found in real networks.

The famous Karate club network analyzed by Zachary is widely used as a benchmark to test the community detection methods [1], [14], [18]. The network consists of 34 members of a karate club as nodes and 78 edges representing friendship between members of the club which was observed over a period of two years. Due to a disagreement between the clubs administrator and its instructor, the club split into two groups. The question we concern is that if we can detect the real groups.

The following parameters are used in MOCD: the population size is 50, the running generation is 100, the crossover ratio is 0.6, and the mutation ratio is 0.4, λ_{str} is 0.5 and λ_{weak} is 0.7. GN and GACD also run on this network. GN finds five communities which are distinguished with the color of the interior of nodes in Fig.4(b). GACD divides the network into 4 groups with the maximal Q value 0.419, which are distinguished with the shape of nodes. They both fail to find the right partition. Fig.4(a) illustrates the 50 non-dominated solutions returned by MOCD. The number of communities of those solutions ranges from 1 to 6. Note that the 50 solutions returned by MOCD actually include the partition results returned by GN and GACD. We label these two solutions with III and I, respectively. Moreover, MOCD successfully reveals the real partition which is denoted by label II in Fig.4(a). In all, MOCD not only finds the community structures discovered by GN and GACD, but also reveals the true structure.

D. Discussion

In the experiments, four networks including the synthetic and social networks are used to validate the effectiveness of MOCD. The optimal solutions returned by MOCD succeed to discover the underlying hierarchical and overlapping structures that hardly can be discovered by one single partition returned by the single-objective approaches. The experiments also show that MOCD can avoid the resolution limit existing in the single-objective approaches (e.g., GN and GACD), because it is able to find small and independent communities. We think the advantages of MOCD are due to the following reason. The real social networks usually are complex and uncertain, because the data of the network are not clean and full of noise. And thus it is nearly impractical to describe the community structure with a fixed partition. MOCD solves the difficulty by providing many solutions in one run. These tradeoff solutions describe the community structure from different angles. A single solution of them may ignore some real structures, but their aggregation (i.e., the Possibility Matrix) can statistically offset the noise and uncertainty and reveal the true and comprehensive information.

V. Conclusion

This paper proposes an multi-objective community detection algorithm consisting of two phases. In the first phase, an multi-objective community detection algorithm is adapted to detect the community structures and return a set of optimal solutions. The second phase, namely model selection phase, we propose three model selection criteria and the Possibility Matrix method that divide the optimal solutions into four parts and

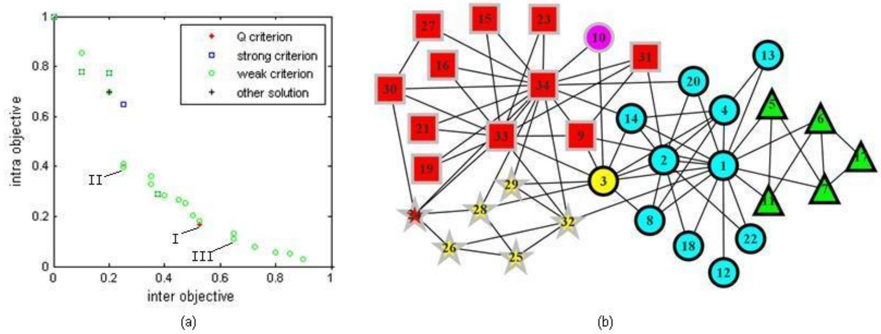

Fig. 4. Multiple solutions of modular structure in Karate network. (a) shows the curve of non-dominated solutions. (b) shows three different partitions. The difference of partitions are made by the color of the boundary of nodes, the color of the interior of nodes, and the shape of nodes, which correspond to the results of real partition & the solution labeled II, GN & the solution labeled III, and GACD & the solution labeled with I, respectively.

assist the DMers to select the preferable ones. Three synthetic and social networks validate the effectiveness of the proposed method (i.e., MOCD). The results show that the MOCD can always find correct partitions by returning a set of optimal solutions. Moreover, with the help of three criteria and the Possibility Matrix, the hierarchical and overlapping nature of the communities can be detected, and DMers can selected a correct solution from different angles.

ACKNOWLEDGMENTS

This work is supported by in part by the National Science Foundation of China (No. 60905025, 61074128, 90924029). It is also supported in part by the National High-tech R&D Program of China (No.2009AA04Z136). Thanks for the contribution of members in TSEG group in BUPT.

REFERENCES

[1] M.E.J.Newman, M.Girvan, "Finding and Evaluating Community Structure in Networks," *Physics Review*, E 2004, 69:026113.
[2] A. Pothen, H. Sinmon, and K-P. Liou, "Partitioning Sparse Matrices with Eigenvectors of Graphs," *SIAM J. Matrix Anal App.*, 1990, 11:430-452.
[3] G.Palla, I.Dereyi, I.Farkas and T.Vicsek, "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society," *Nature*, 2005, 435(7043): 814-818.
[4] L.Danon, A.Diaaz-Guilera, J.Duch and A.Arenas, "Comparing Community Structure Identification," *Journal of Statistical Mechanics: Theory and Experiments*, 2005.
[5] R.Martin and T. B.Carl, "An information-theoretic framework for resolving community structure in complex networks," *PNAS*, 2007, 104, 7327-7331.
[6] S.Fortunato and M.Barthelemy, "Resolution Limit in Community Detection," *Proceedings of the National Academy of Sciences*, 2007,104(1):36-41.
[7] J.Reichardt and S.Bornholdt, "Statistical Mechanics of Community Detection," *Physics Review* E, 2006,74(1):016110.
[8] A.Arenas, A.Fernandez and A. Gomez, "Multiple Resolution of Modular Structure of Complex Networks," *arXiv:physics/0703218v1*, 2007.
[9] J.M.Kumpula, J.Saramaki, K.Kaski and J.Kertesz, "Limit Resolution and Multiresolution Models in Complex Network Community Detection," *arXiv:0706. 2230v2*, 25 Jan 2008.
[10] C.L.Hwang and A.S.M.Masud, "Multiple objective Decision Making-Methods and Applications", *Springer Verlag, Berlin, Germany*, 1979.
[11] F.Radicchi, C.Castellano, F.Cecconi, V.Loreto, and D.Parisi, "Defining and Identifying Communities in Networks," *PNAS*, 2004, 101(9):2658-2663.
[12] A.Clauset, M.E.J.Newman, and C.Moore, "Finding community structure in very large networks", *Physical Review* E, 2004, vol 70:06611.
[13] U.Brandes, D.Delling, M.Gaetler, et al, "On Modularity Clustering", *IEEE Transactions on Knowledge and Data Engineering*, 2008,20(2):172-188.
[14] J.Duch, A.Arenas, "Community Detection in Complex Networks using Extremal Optimization," *arXiv:condmat* /0501368, 2005.
[15] M.Tasgin and H.Bingol, "Community Detection in Complex Networks using Genetic Algorithm," *arXiv:cond-mat* /0604419, 2006.
[16] K.Deb, "Multiobjective Optimization using Evolutionary Algorithms," *U.K: Wiley*, 2001.
[17] D.W.Corne, N.R.Jerram, J.D.Knowles, and M.J.Oates, "PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization," in *Proc. Genetic Evol. Comput. Conf.* 2001, pp.283-290.
[18] C. Shi, Y. Wang, B. Wu, C. Zhong, "A New Genetic Algorithm for Community Detection," *Complex Sciences*, 2009, 5(1):1298-1309.
[19] C.Pizzuti, "GA-Net: a genetic algorithm for community detection in social networks," *in PPSN2008*, pp.1081-1090.
[20] C.Pizzuti, "Community Detection in Social Networks with Genetic Algorithms", *GECCO2008*, 2008:11371138.
[21] C. Shi, C. Zhong, Zhenyu Yan, et al., "A Multi-Objective Optimization Approach for CommunityDetection," *CEC2010*
[22] C. Pizzuti, "A Multi-objective Genetic Algorithm for Community Detection in Networks," *ICTAI09* 379-386. http://www-personal.umich.edu/ mejn/netdata.