Detecting Community Structure of Complex Networks by Simulated Annealing with Optimal Prediction

Jian Liu
LMAM and School of Mathematical Sciences, Peking University
Beijing 100871, P.R. China
Email: dugujian1984@sina.com

Na Wang
Beijing University of Posts and Telecommunications
Beijing 100876, P.R. China
Email: nawang007@163.com

***Abstract*—Given a large and complex network, we would like to find the best partition of this network into a small number of clusters. This question has been addressed in many different ways. Here we utilize the simulated annealing strategy to maximize the modularity of a network with our previous hard partitioning formulation for the community structure, which is based on the optimal prediction of a random walker Markovian dynamics on the network. It is demonstrated that this simulated annealing with optimal prediction (SAOP) algorithm can efficiently and automatically determine the number of communities during the cooling procedure associated with iterative steps. Moreover, the algorithm is successfully applied to three model problems.**

***Keywords*-complex networks; community structure; optimal prediction; modularity; simulated annealing.**

I. Introduction

In recent years we have seen an explosive growth of interest and activity on the structure and dynamics of complex networks [1], [2], [3]. This is partly due to the influx of new ideas, particularly ideas from statistical mechanics, to the subject, and partly due to the emergence of interesting and challenging new examples of complex networks such as the internet and wireless communication networks. Network models have also become popular tools in social science, economics, the design of transportation and communication systems, banking systems, powergrid, etc, due to our increased capability of analyzing these models. Since these networks are typically very complex, it is of great interest to see whether they can be reduced to much simpler systems. In particular, much effort has gone into partitioning the network into a small number of clusters [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], which are constructed from different viewing angles comparing different proposals in the literature.

In a previous paper [13], the authors proposed an approach to partition the networks based on optimal prediction theory proposed by Chorin and coworkers [15], [16]. The basic idea is to associate the network with the random walker Markovian dynamics [17], then introduce a metric on the space of Markov chains (stochastic matrices), and optimally reduce the chain under this metric. The final minimization problem is solved by an analogy to the traditional k-means algorithm in clustering analysis [18]. This approach also bears some similarity to the MNCut algorithms in image segmentation [4], [5] and the diffusion maps in data mining [12].

In traditional clustering literature, a function called validity index [19], [20], [21], [22] is often used to evaluate the quality of clustering results. The optimal number of clusters can be determined by selecting the minimum value of the index. With the same idea, the most widely used validation in detecting communities of networks is modularity [7], [10], [11], which has larger values indicating stronger community structure. Then simulated annealing strategy [23], [24], [25] is utilized to obtain the maximum value of modularity. This kind of simulated annealing operated with iteration methods is quite different with the previous work [26], [27], since the process of iteration can accelerate the tendency in maximizing the modularity.

We constructed our algorithm — simulated annealing with optimal prediction (SAOP) for network partitioning. From the numerical performance to three model problems: the ad hoc network with 128 nodes, sample networks generated from Gaussian mixture model and karate club network, we can see that our algorithm can efficiently and automatically determine the number of communities during the cooling procedure associated with iterative steps.

The rest of the paper is organized as follows. In Section II, we briefly introduce the framework of optimal partition algorithm [13] and the definition of modularity in networks. After reviewing the idea of simulated annealing, we proposed our algorithm (SAOP) and the corresponding strategies in Section III. In Section IV, we apply SAOP to three model problems mentioned before. The partitioning results are typically compared in detail. Finally we make the conclusion in Section V.

II. Problem Formulations

A. Network Partition Based on Optimal Prediction

In [13], a new strategy for reducing the random walker Markovian dynamics based on optimal prediction theory [15],

978-1-4244-4507-3/09/$25.00 ©2009 IEEE

[16] is proposed. Let $G(S, E)$ be a network with n nodes and m edges, where S is the nodes set, $E = \{e(x, y)\}_{x,y\in S}$ is the weight matrix and $e(x, y)$ is the weight for the edge connecting the nodes x and y. We can relate this network to a discrete-time Markov chain with stochastic matrix P with entries $p(x, y)$ given by

$$p(x, y) = \frac{e(x, y)}{d(x)}, \qquad d(x) = \sum_{z\in S} e(x, z), \tag{1}$$

where $d(x)$ is the degree of the node x [12], [17], [28]. This Markov chain has stationary distribution

$$\mu(x) = \frac{d(x)}{\sum_{z\in S} d(z)} \tag{2}$$

and it satisfies the detailed balance condition [13].

The basic idea in [13] is to introduce a metric for the stochastic matrix $p(x, y)$

$$\|p\|_\mu^2 = \sum_{x,y\in S} \frac{\mu(x)}{\mu(y)} |p(x, y)|^2 \tag{3}$$

and find the reduced Markov chain $\tilde{p}$ by minimizing the distance $\|\tilde{p} - p\|_\mu$. For a given partition of S as $S = \cup_{k=1}^N S_k$ with $S_k \cap S_l = \emptyset$ if $k \neq l$, let $\hat{p}_{kl}$ be the coarse grained transition probability from S_k to S_l on the state space $\mathbb{S} = \{S_1, \ldots, S_N\}$. This matrix can be naturally lifted to the space of stochastic matrices on the original state space S via

$$\tilde{p}(x, y) = \sum_{k,l=1}^N \mathbf{1}_{S_k}(x)\hat{p}_{kl}\mu_l(y), \tag{4}$$

where $\mathbf{1}_{S_k}(x) = 1$ if $x \in S_k$ and $\mathbf{1}_{S_k}(x) = 0$ otherwise, and

$$\mu_k(x) = \frac{\mu(x)\mathbf{1}_{S_k}(x)}{\hat{\mu}_k}, \qquad \hat{\mu}_k = \sum_{z\in S_k} \mu(z). \tag{5}$$

Based upon this formulation, we can find the optimal $\hat{p}_{kl}$ for any fixed partition. With this optimal form $\hat{p}_{kl}$, we further search for the best partition $\{S_1, \cdots, S_N\}$ with the given number of communities N by minimizing the optimal prediction error $J = \|\tilde{p} - p\|_\mu^2$. This is the theoretical basis for constructing the k-means algorithm for the community structure of complex networks in [13], which is considered to address this optimization issue which guarantees convergence towards a local minimum

(1) Initialize the partition $\{S_k^{(0)}\}_{k=1}^N$ at random in the diffusion space;

(2) For $n \geq 0$, compute $\hat{p}_{kl}^{(n)}$ according to

$$\hat{p}_{kl}^{(n)} = \sum_{x\in S_k^{(n)}, y\in S_l^{(n)}} \mu_k(x)p(x, y); \tag{6}$$

(3) For $n \geq 0$, update the partition using

$$S_k^{(n+1)} = \Big\{x : k = \arg\min_l D(x, S_l^{(n)})\Big\}, \quad k = 1, \cdots, N, \tag{7}$$

where

$$D(x, S_k) = \sum_{l=1}^N \sum_{y\in S_l} \mu(x)\mu(y) \left(\frac{p(x, y)}{\mu(y)} - \frac{\hat{p}_{kl}}{\hat{\mu}_l}\right)^2; \tag{8}$$

(4) Repeat (2) to (3) until convergence.

B. Modularity of Complex Networks

A measure of the quality of a particular partition of a network, which we call the modularity, which is based on a previous measure of assortative mixing proposed by Newman [29], defined by

$$\begin{aligned} Q \quad = \quad & (\text{number of edges within communities}) \\ & -(\text{expected number of such edges}). \end{aligned} \tag{9}$$

It is a function of the particular partition of the network into groups, with larger values indicating stronger community structure. Some methods are proposed to find good partitions of a network into communities by optimizing the modularity over possible divisions [7], [29], which has proven highly effective in practice [27]. Our work is quite different that we using simulated annealing to find the maximum of Q through a iterative procedure as k-means algorithm above. Let $p^e(x, y)$ be the probability for an edge to fall between every pair of nodes x and y. More precisely, $p^e(x, y)$ is the expected number of edges between x and y, a definition that allows for the possibility that there may be more than one edge between a pair of vertices, which happens in certain types of networks. For a given partition $\{S_i\}_{i=1}^N$, the modularity can be written in the following form

$$Q = \frac{1}{m} \sum_{i=1}^N \sum_{x,y\in S_i} \Big(e(x, y) - p^e(x, y)\Big), \tag{10}$$

where m is the number of edges in the network, and

$$p^e(x, y) = \frac{d(x)d(y)}{2m}, \tag{11}$$

which can be obtained after some simple manipulation [10].

III. The Algorithms

A. Simulated annealing

The first simulated annealing algorithm was proposed by Metropolis et al. in 1953 [23], which was motivated by simulating the physical process of annealing solids. The process can be described as follows. Firstly, a solid is heated from a high temperature and then cooled slowly so that the system at any time is approximately in thermodynamic equilibrium. At equilibrium, there may be many configurations with each one corresponding to a specific energy level. The chance of accepting a change from the current configuration to a new configuration is related to the difference in energy between the two states. Kirkpatrick et al. were the first to introduce simulated annealing to optimization problems in 1982 [24]. Kein and Dubes [25] have developed algorithms based on simulated annealing to find the global minimum solution using fuzzy c-means and other crisp clustering methods. However,

the number of clusters has to be declared in advance for all of these techniques.

Let $E = -Q$. $E^{(n)}$and $E^{(n+1)}$ represent the current energy and new energy respectively. $E^{(n+1)}$is always accepted if it satisfies $E^{(n+1)} < E^{(n)}$, but if $E^{(n+1)} > E^{(n)}$ the new energy level is only accepted with a probability as specified by $\exp(-\frac{1}{T}\triangle E^{(n)})$, where $\triangle E^{(n)} = E^{(n+1)} - E^{(n)}$ is the difference of energy and T is the current temperature. Worse solutions are accepted based on the change in solution quality which allows the search to avoid becoming trapped at local minima. The temperature is then decreased gradually and the annealing process is repeated until no more improvement is reached or any termination criteria have been met.

B. The algorithmic aspects and corresponding strategies

At a given temperature, the new state is accepted with a probability $\exp(-\frac{1}{T}\triangle E^{(n)})$, where the energy, equal to minus modularity, is used to evaluate a partition. The initial state is generated by randomly N clusters, where N is a integer within the range $[N_{\min}, N_{\max}]$. We usually choose values $N_{\min} = 2$ and $N_{\max} = \sqrt{n}$. The initial temperature T is set to a high temperature $T_{\max}$. A neighbor of the current state is produced by randomly flipping one spin, then the energy of the new state is calculated. The new state is kept if the acceptance requirement is satisfied. This process will be repeated for R times at the given temperature. A cooling rate $0 < \alpha < 1$ decreased the current temperature until reached the bound $T_{\min}$. The whole procedure of the simulated annealing with optimal prediction (SAOP) algorithm is summarized as follows

(1) Set parameters $T_{\max}$, $T_{\min}$, $N_{\min}$, $N_{\max}$, α and R. Choose N randomly within range $[N_{\min}, N_{\max}]$ and initialize the partition $\{S_k^{(0)}\}_{k=1}^{N}$ randomly. Set the current temperature $T = T_{\max}$.

(2) Compute the corresponding $\hat{p}_{kl}^{(0)}$ according to (6) and calculate the initial energy $E^{(0)}$ using the definition of modularity (10); Set $n^* = 0$.

(3) For $n = 0, 1, \cdots, R$, do the following

(3.1) Generate a new partition $\{S_k^{(n)}\}_{k=1}^{N'}$ according to our proposal below and set $N = N'$;

(3.2) Update the corresponding $\hat{p}_{kl}^{(n)}$ according to (6);

(3.3) Update the partition $\{S_k^{(n+1)}\}_{k=1}^{N}$ using (7) and calculate the new energy $E^{(n+1)}$ using (10);

(3.4) Accept or reject the new state. If $E^{(n+1)} < E^{(n)}$ or $E^{(n+1)} > E^{(n)}$ with $u \sim \mathcal{U}[0,1]$, $u < \exp\{-\frac{1}{T}\triangle E^{(n)}\}$, then accept the new solution by setting $n = n + 1$. Else,reject it;

(3.5) Update the optimal state. If $E^{(n)} < E^{(n^*)}$, set $n* = n$.

(4) Cooling temperature $T = \alpha \cdot T$. If $T < T_{\min}$, go to Step (5); Else, set $n = n^*$, repeat Step (3).

(5) Output the optimal solution $\{S_k^{(n^*)}\}_{k=1}^{N}$ and the minimum energy $E^{(n^*)}$ of the whole procedure.

Our proposal to the process of generating a set of new partitions in Step (3.1) comprises three functions, which are deleting a current community, splitting a current community and perturbing a current community. At each iteration, one of the three functions can by randomly chosen and the community strength $M_k = \hat{p}_{kk}, k = 1, \cdots, N$ is used to select a community, which reflect the possibility of community S_k remains in itself but not tend flowing into others. Obviously, the structure of a community is more strong if its strength is more larger. The three functions are described below

- Delete Community. The cluster with the minimum community strength S_d is identified, which should be deleted from the current partition with merging the nodes in S_d to S_k, here $k = \arg\max_m \hat{p}_{dm}$.
- Split Community. The cluster with the maximum community strength S_s is chosen, which is then replaced by two new communities if the minimum stationary distribution $\min_{x \in S_s} \mu(x)$ is much smaller than the average stationary distribution. We obtain the new community S_{N+1} by randomly choosing half of the nodes in S_s.
- Perturb Community. As the same to Delete Community, we choose the cluster with the minimum community strength S_p to perturb. Let n_p be the number of nodes in S_p, n_k in S_k and n_l in S_l, here $k = \arg\max_m \hat{p}_{dm}$ and $l = \arg\max_m \hat{p}_{mp}$. Generate $r \sim \mathcal{U}[-1, 1]$. If $r > 0$, randomly choose $r \cdot \eta \cdot n_p$ nodes in S_p classified to S_k; Else, randomly choose $-r \cdot \eta \cdot n_l$ nodes in S_l classified to S_k. Here $\eta \in (0, 1]$ is some given parameter.

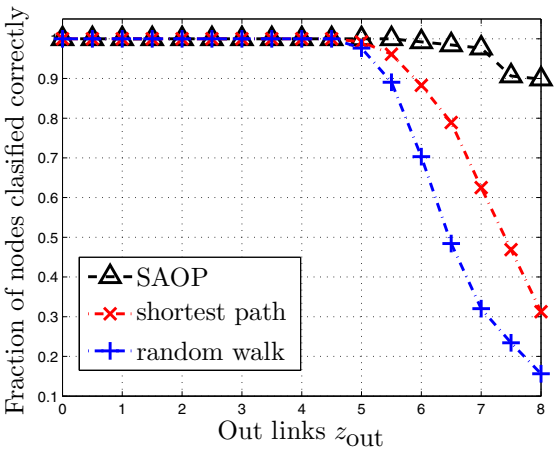


Fig. 1. The fraction of nodes classified correctly of ad hoc network by SAOP and the methods used in [7]. It seems that SAOP has better partition result than shortest path and random walk methods [7].

IV. Experimental results

A. Ad hoc network with 128 nodes

We apply our methods to the ad hoc network with 128 nodes in this subsection. The ad hoc network is a typical benchmark problem considered in many papers [7], [8], [13], [27]. It has a known community structure and is constructed as follows. Suppose we choose $n = 128$ nodes, split into 4 communities containing 32 nodes each. Assume pairs of nodes belonging to the same communities are linked with probability p_{in}, and pairs belonging to different communities with probability p_{out}.

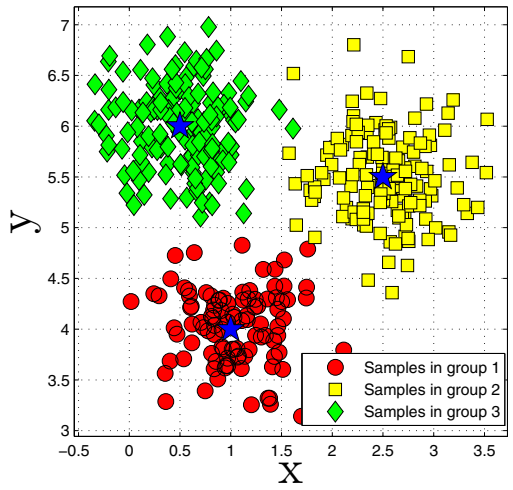


Fig. 2. 40 sample points generated from the given 3-Gaussian mixture distribution. The star symbols represent the centers of each Gaussian component. The circle, square and diamond shaped symbols represent the position of sample points in each component respectively.

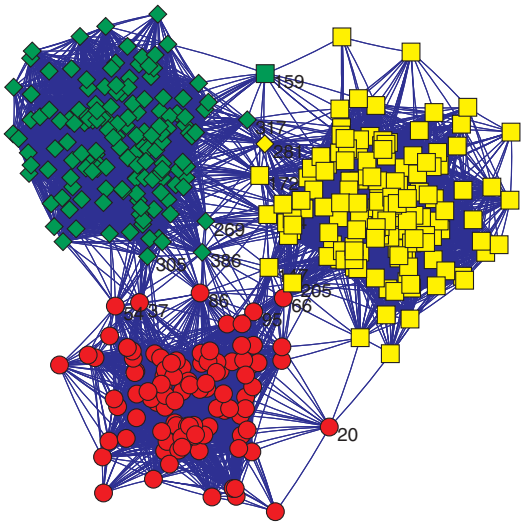


Fig. 3. Partition of the network generated from the sample points in Figure 2 with the parameter $dist = 0.8$ by SAOP.

These values are chosen so that the average node degree, d, is fixed at $d = 160$. In other words p_{in} and p_{out} are related as

$$31p_{\text{in}} + 96p_{\text{out}} = 16. \tag{12}$$

Here we naturally choose the nodes group $S_1 = \{1 : 32\}, S_2 = \{33 : 64\}, S_3 = \{65 : 96\}, S_4 = \{97 : 128\}$.

We change z_{out} from 0 to 8 and look into the fraction of nodes which correctly classified. By setting $T_{\max} = 3$, $T_{\min} = 10^{-3}$, $\alpha = 0.9$, $R = 50$, $\eta = 0.005$, we make clustering by SAOP. The fraction of correctly identified nodes is shown in Figure 1, comparing with the two methods described in [7]. SAOP perform noticeably better than the two previous methods, especially for the more difficult cases when z_{out} is large.

B. Sample Network Generated from Gaussian Mixture Model

To further test the validity of the algorithms, we apply them to a sample network generated from a Gaussian mixture model. This model is quite related the concept random geometric graph proposed by Penrose [30] except that we take Gaussian mixture here compared with uniform distribution in [30].

We generate n sample points $\{\boldsymbol{x}_i\}$ in two dimensional Euclidean space subject to a K-Gaussian mixture distribution at first

$$\sum_{i=1}^{K} q_i G\left(\boldsymbol{\mu}_i, \Sigma_i\right), \tag{13}$$

where $\{q_i\}$ are mixture proportions satisfying $0 < q_i < 1$, $\sum_{i=1}^{K} q_i = 1$. $\boldsymbol{\mu}_i$ and Σ_i are the mean positions and covariance matrices for each component, respectively. Then we generate the network with a thresholding strategy. That is, if $|\boldsymbol{x}_i - \boldsymbol{x}_j| \leq dist$, we set an edge between the i-th and j-th node; otherwise they are not connected. With this strategy, the topology of the network is induced by the metric. As a consequence, some properties of the network, say the clustering nature, may be inherited from the case with metric. This is our basic motivation with this model.

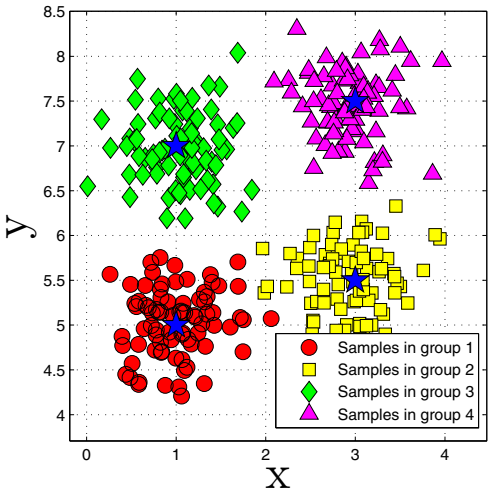


Fig. 4. 320 sample points generated from the given 4-Gaussian mixture distribution. The star symbols represent the centers of each Gaussian component. The circle, square, diamond and triangle shaped symbols represent the position of sample points in each component respectively.

First we take $n = 400$ and $K = 3$, then generate the sample points with the means

$$\boldsymbol{\mu}_1 = (1.0, 4.0)^T, \boldsymbol{\mu}_2 = (2.5, 5.5)^T, \boldsymbol{\mu}_3 = (0.5, 6.0)^T, \tag{14a}$$

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{pmatrix} 0.15 & 0 \\ 0 & 0.15 \end{pmatrix}. \tag{14b}$$

Here we pick nodes 1:100 in group 1, nodes 101:250 in group

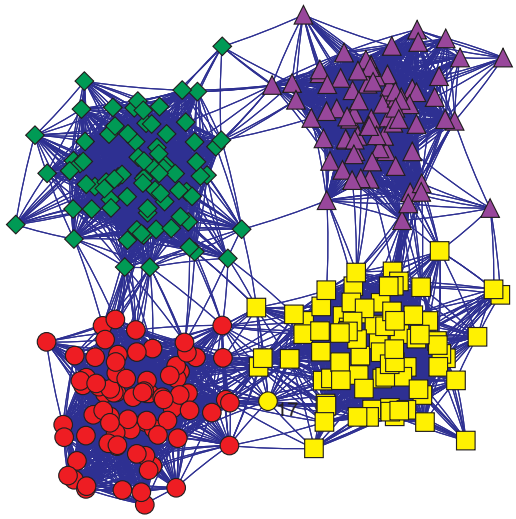


Fig. 5. Partition of the network generated from the sample points in Figure 4 with the parameter $dist = 0.8$ by SAOP.

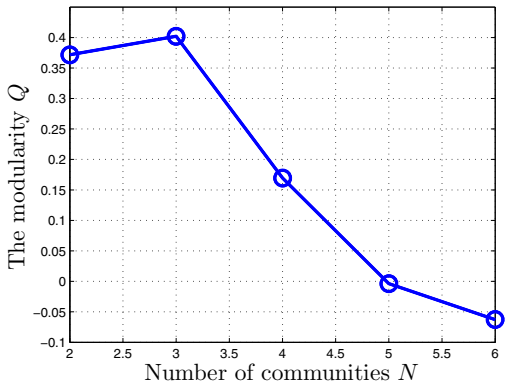


Fig. 6. The modularity (10) changed with the number of communities N in karate club network by optimal prediction algorithm [13]. The optimal community structure was reached at $N = 3$ and the corresponding modularity was $Q = 0.4020$.

2 and nodes 251:400 in group 3 for simplicity. With this choice, approximately $q_1 = 100/400, q_2 = q_3 = 150/400$. We take $dist = 0.8$ in this example. The sample points are shown in Figure 2. By setting $T_{\max} = 3$, $T_{\min} = 10^{-2}$, $\alpha = 0.9$, $R = 20$, $\eta = 0.005$, we obtain $N = 3$ and the corresponding $Q = 0.6241$. The partition result is shown in Figure 3.

Next we take $n = 320$ and $K = 4$, where nodes 1:80 are in group 1, nodes 81:160 in group 2, nodes 161:240 in group 3 and nodes 241:320 in group 4. This means approximately $q_1 = q_2 = q_3 = q_4 = 80/320$. The other model parameters are chosen as

$$\begin{aligned} \boldsymbol{\mu}_1 &= (1.0, 5.0)^T, \boldsymbol{\mu}_2 = (3.0, 5.5)^T, \\ \boldsymbol{\mu}_3 &= (1.0, 7.0)^T, \boldsymbol{\mu}_4 = (3.0, 7.5)^T, \end{aligned} \tag{15a}$$

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \begin{pmatrix} 0.15 & 0 \\ 0 & 0.15 \end{pmatrix}. \tag{15b}$$

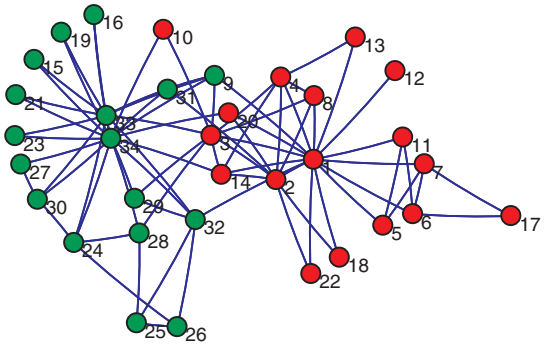
Here we take $dist = 0.8$. The sample points are shown in Figure 4. Then we generate the network and make clustering with our methods. By setting $T_{\max} = 3$, $T_{\min} = 10^{-2}$, $\alpha = 0.9$, $R = 20$, $\eta = 0.005$, we obtain $N = 4$ and the corresponding $Q = 0.7301$. The partition result is shown in Figure 5. We can see that the results are reasonable and our algorithm goes smoothly with several hundreds of nodes.

C. Karate Club Network

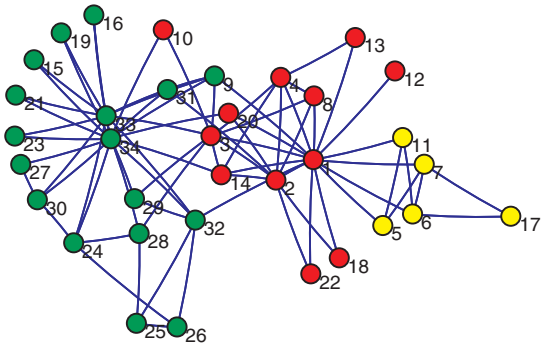
This network was constructed by Wayne Zachary after he observed social interactions between members of a karate club at an American university [31]. Soon after, a dispute arose between the clubs administrator and main teacher and the club split into two smaller clubs. It has been used in several papers to test the algorithms for finding community structure in networks [7], [9], [10], [11], [13]. There are 34 nodes in karate club network (see Figures 7 and 8), where each node represents one member in the club. In Zachary's original partition, each node belongs to only one sub-club after splitting. We label it as red or green color in Figure 7(a), where only node 10 is misclassified, to show its attribute in the graph representation. When the annealing strategy were not used, i.e. only the optimal prediction algorithm [13] was operated, the modularity Q changed with the number of communities N is shown in Figure 6 and the optimal community structure in Figure 7. We can see the optimal community structure was reached at $N = 3$ and the corresponding modularity was $Q = 0.4020$. Then we utilized SAOP by setting $T_{\max} = 3$, $T_{\min} = 10^{-3}$, $\alpha = 0.9$, $R = 50$, $\eta = 0.005$, we obtain $N = 4$ and the corresponding $Q = 0.4198$. The partition result is shown in Figure 8. We can see that SAOP find a better community structure of the karate club network than the case when only the optimal prediction algorithm [13] was operated, according to the key quality of modularity. The result in Figure 8 also seem reasonable to the network structure itself when the real-world information is removed.

V. Conclusions

We utilize simulated annealing strategy to search the global maximum of the modularity of a given network, associating with the previous optimal prediction algorithm [13] for the network partitioning problem in this paper. The proposed algrorithm — simulated annealing with optimal prediction (SAOP), is constructed and succeed in three model problems, including the ad hoc network with 128 nodes, the sample network generated from Gaussian mixture model and the karate club network. It is demonstrated by experiments that our algorithm can not only find the optimal community structure, but also determine the number of communities efficiently and automatically during the cooling procedure associated with iterations. This leads to the automatical model selection problem which people are sometimes interested in [2], [7], [9], [11], [27].



(a)



(b)

Fig. 7. Partition of the karate club network by optimal prediction algorithm [13] in the case $N = 2$ and $N = 3$ respectively. Note that node 10 is misclassified in both cases. (a) $N = 2$; (b)$N = 3$.

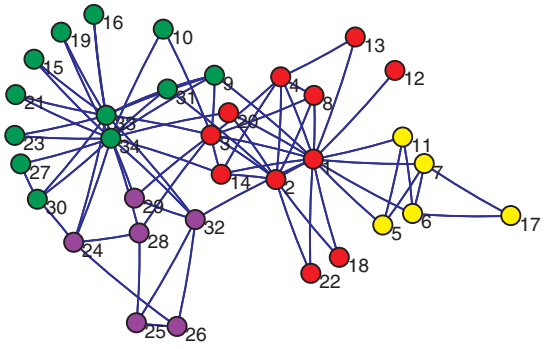


Fig. 8. Partition of the karate club network by SAOP. The corresponding modularity $Q = 0.4198$. Note that node 10 is classified correctly this time.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China under grant 10871010 and the National Basic Research Program of China under grant 2005CB321704. The authors thank Prof. Tiejun Li for many discussions.

REFERENCES

[1] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[2] M. Newman, A.-L. Barabási, and D. J. Watts, *The structure and dynamics of networks*. Princeton: Princeton University Press, 2005.

[3] N. R. Council, *Network Science. National Academy of Sciences*, Washington DC, 2005.

[4] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern. Anal. Mach. Intel.*, vol. 22, no. 8, pp. 888–905, 2000.

[5] M. Meilă and J. Shi, "A random walks view of spectral segmentation," in *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, 2001, pp. 92–97.

[6] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[7] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, 2004.

[8] M. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 6, p. 066133, 2004.

[9] ——, "Detecting community structure in networks," *Eur. Phys. J. B*, vol. 38, no. 2, pp. 321–330, 2004.

[10] ——, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, p. 036104, 2006.

[11] ——, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006.

[12] S. Lafon and A. Lee, "Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization," *IEEE Trans. Pattern. Anal. Mach. Intel.*, pp. 1393–1403, 2006.

[13] W. E, T. Li, and E. Vanden-Eijnden, "Optimal partition and effective dynamics of complex networks," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 23, pp. 7907–7912, 2008.

[14] J. M. Hofman and C. H. Wiggins, "A bayesian approach to network modularity," *Phys. Rev. Lett.*, vol. 100, p. 258701, 2008.

[15] A. J. Chorin, A. P. Kast, and R. Kupferman, "Unresolved computation and optimal predictions," *Comm. Pure Appl. Math.*, vol. 52, no. 10, pp. 1231–1254, 1999.

[16] A. J. Chorin, "Conditional expectations and renormalization," *Multi. Model. Simul.*, vol. 1, p. 105, 2003.

[17] L. Lovasz, "Random walks on graphs: A survey," *Combinatorics, Paul Erdos is Eighty*, vol. 2, pp. 1–46, 1993.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

[19] J. C. Bezdek, "Cluster validity with fuzzy sets," *Cybernetics and Systems*, vol. 3, no. 3, pp. 58–73, 1973.

[20] ——, "Mathematical models for systematics and taxonomy," in *Proc. 8th Int. Conf. on Numerical Taxanomy*, 1975, pp. 143–166.

[21] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Tran. Pattern Anal. Mach. Intel.*, vol. 13, no. 8, pp. 841–847, 1991.

[22] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995.

[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, p. 1087, 1953.

[24] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[25] R. W. Klein and R. C. Dubes, "Experiments in projection and clustering by simulated annealing," *Pattern Recognition*, vol. 22, no. 2, pp. 213–220, 1989.

[26] C. P. Massen and J. P. K. Doye, "Identifying communities within energy landscapes," *Phys. Rev. E*, vol. 71, no. 4, p. 046101, 2005.

[27] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Stat. Mech.*, vol. 9, p. P09008, 2005.

[28] F. Chung, *Spectral Graph Theory. American Mathematical Society*, Rhode Island, 1997.

[29] M. Newman, "Mixing patterns in networks," *Phys. Rev. E*, vol. 67, no. 2, p. 26126, 2003.

[30] M. Penrose, *Random Geometric Graphs*. Oxford: Oxford University Press, 2003.

[31] W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthrop. Res.*, vol. 33, no. 4, pp. 452–473, 1977.