Dynamic evolutionary community detection algorithms based on the modularity matrix*

Chen Jian-Rui(陈建芮)†, Hong Zhi-Min(洪志敏), Wang Li-Na(汪丽娜), and Wu Lan(乌 兰)

College of Science, Inner Mongolia University of Technology, Hohhot 010051, *China*

(Received 31 January 2014; revised manuscript received 1 June 2014; published online 23 September 2014)

Motivated by the relationship of the dynamic behaviors and network structure, in this paper, we present two efficient dynamic community detection algorithms. The phases of the nodes in the network can evolve according to our proposed differential equations. In each iteration, the phases of the nodes are controlled by several parameters. It is found that the phases of the nodes are ultimately clustered into several communities after a short period of evolution. They can be adopted to detect the communities successfully. The second differential equation can dynamically adjust several parameters, so it can obtain satisfactory detection results. Simulations on some test networks have verified the efficiency of the presented algorithms.

Keywords: community detection, dynamic evolutionary, modularity matrix, synchronization

PACS: 89.75.Fb, 89.75.Hc, 84.30.Bv **DOI:** 10.1088/1674-1056/23/11/118903

1. Introduction

Complex networks are a useful tool to model many real systems, such as social networks, food webs, the world-wide-web, and biological organizations.[1] Dynamic behaviors[2–4] and topological structure analysis[5–9] are more active among the research of complex networks. Modular structures constitute one of the most important properties of real-world networked systems. A network can be represented as a graph composed by a large number of edges connecting vertices within individual groups, with only a low concentration of edges between these groups.[10] A single community may represent different functional units of the system, and it plays a very important role for understanding the entire network and analyzing network characteristics.

Recently, many community detection algorithms are presented based on various approaches such as edge clustering coefficients,[11] dissimilarity and similarity measures,[12,13] spectral analysis methods,[14,15] physical models,[16,17] and optimization.[18,19] Moreover, many dynamic clustering algorithms based on synchronization are proposed to discover the community in the networks.[20–23] They have shown that the dynamical synchronization process in complex networks is a suitable method to reveal the different topological structure. Most of these dynamic algorithms are based on the Kuramoto model[24] or on its variants such as the opinion changing rate model.[25] In Ref. [26], they utilized a proper weighting scheme and an algorithm based on the partial synchronization phenomenon as pre-processing and post-processing steps to improve the Newman–Fast algorithm. A parameterless mechanism adapts the characteristic frequencies of coupled oscillators according to a dynamic connectivity matrix deduced from correlated data.[27] In Ref. [28], a generalized Kuramoto model is considered, the phases of the nodes are naturally separated into several clusters after a period of evolution, and each cluster corresponds to a community in the network.

Based on the above work, in this paper, we adopt synchronization dynamic evolutionary behaviors to reveal the communities in the networks. Firstly, we would improve the Kuramoto model for the purpose of community detection based on the modularity matrix. Secondly, we further generalize the model to enhance the synchronization in the community. The phases of the nodes will dynamically evolve according to the two algorithms. The phases of the nodes in the same community would cluster closely and the phases of the nodes in the different communities would move far away. The community structures can be quickly detected.

The rest of this paper is organized as follows. The dynamic community detection models and two algorithms DE_M and DC_M$^{\alpha}$ are presented in Section 2. The simulation results are given in Section 3 and Section 4 concludes the paper.

2. Network models and community detection algorithms

2.1. Dynamic evolutionary algorithm based on modularity matrix

The following generalized Kuramato model has been proposed:[28]

$$\dot{\theta}_i = \omega_i + \frac{K_{\rm p}}{N}\sum_{j=1}^{N} a_{ij}\sin(\theta_j - \theta_i)$$

*Project supported by the National Natural Science Foundation of China (Grant No. 61272279), the TianYuan Special Funds of the National Natural Science Foundation of China (Grant No. 11326239), the Higher School Science and Technology Research Project of Inner Mongolia, China (Grant No. NJZY13119), and the Inner Mongolia University of Technology, China (Grant No. ZD201221).

†Corresponding author. E-mail: jianrui_chen@sina.com

© 2014 Chinese Physical Society and IOP Publishing Ltd

$$+\frac{K_{\rm n}}{N}\sum_{j=1}^{N}(1-a_{ij})\sin(\theta_j-\theta_i). \quad (1)$$

If the j-th node is connected with the i-th node, $a_{ij}=1$, otherwise, $a_{ij}=0$.

Furthermore, it can be rewritten as

$$\dot{\theta}_i=\omega_i+\frac{K_{\rm p}}{N}\sum_{j\in N_i}\sin(\theta_j-\theta_i)+\frac{K_{\rm n}}{N}\sum_{j\notin N_i}\sin(\theta_j-\theta_i), \quad (2)$$

where N_i is the set of neighbors of the i-th node. We set $K_{\rm p}>0$ to make the phases of two connected oscillators in the network evolve together, and $K_{\rm n}\leq 0$ to make the phases of two unconnected oscillators in the network evolve far away. They have realized community detection successfully. Generally, different neighbors of the node have different close relationships. The probability with close relationship nodes in the same community is larger. For example, in social networks, the probability of good friends being together in a community is larger than the ordinary friends. In this paper, the similarity of the nodes is adopted to describe the close relationship between them. Equation (2) would be improved to obtain better community detection performance based on this similarity. Newman introduced modularity to evaluate the goodness of a particular division of a network into communities.[29]

We consider such a network, in which there are N nodes and m edges, represented by an adjacent matrix $\boldsymbol{A}$, whose element $a_{ij}=1$ if there is a link between nodes i and j, and $a_{ij}=0$ otherwise. Modularity is defined as

$$Q=\frac{1}{2m}\sum_{i,j\in V}\left(a_{ij}-\frac{k_ik_j}{2m}\right)\delta(c_i,c_j), \quad (3)$$

where k_i and k_j are the degrees of vertices i and j, respectively; $\delta(c_i,c_j)$ is the Kronecker symbol equal to 1 if the communities c_i and c_j, to which i and j belong, are the same, and to 0 otherwise. A new real symmetric matrix $\boldsymbol{B}$ with elements

$$B_{ij}=a_{ij}-\frac{k_ik_j}{2m} \quad (4)$$

are called the modularity matrix.[30]

Here, it can be rewritten as follows:

$$B_{ij}=\begin{cases}1-\dfrac{k_ik_j}{2m}, & a_{ij}=1,\\ -\dfrac{k_ik_j}{2m}, & a_{ij}=0.\end{cases} \quad (5)$$

Generally, if the j-th node is the neighbor of the i-th node, $B_{ij}>0$; otherwise, $B_{ij}<0$. However, we discover B_{ij} may be less than zero when the j-th node is the neighbor of the i-th node, i.e., $k_i/k_j2m>1$. For example, in Zachary's karate club network, $k_1=16$, $k_3=10$, $m=78$, and thus $B_{13}=-0.0256<0$. To guarantee the symbol of the neighbors of the same node is higher than zero, we define the similarity of the two nodes as follows:

$$S_{ij}=\begin{cases}B_{ij}>0, & a_{ij}=1 \quad \text{and} \quad B_{ij}>0,\\ \dfrac{2(a_{ij}+\sum_{k=1}^{N}a_{ik}a_{jk})}{k_i+k_j}>0, & a_{ij}=1 \quad \text{and} \quad B_{ij}\leq 0,\\ B_{ij}<0, & a_{ij}=0.\end{cases} \quad (6)$$

The term $2(a_{ij}+\sum_{k=1}^{N}a_{ik}a_{jk})/(k_i+k_j)$ is motivated by Ref. [26]. $\sum_{k=1}^{N}a_{ik}a_{jk}$ contains the information of their common neighbors ($a_{ik}a_{jk}=1$ if and only if $a_{ik}=1$ and $a_{jk}=1$, while $a_{ik}a_{jk}=0$ in other cases). The quantity $k_ik_j/2m$ is the expected number of edges between vertices i and j in a null model where edges are placed at random, while the distribution of degrees remains the same.

Thus, our dynamic evolutionary algorithm based on modularity matrix (denoted as DE_M) is executed by the following differential equation:

$$\begin{aligned}\dot{\theta}_i &= \omega_i+\frac{K_1}{N}\sum_{j\in N_i}S_{ij}\sin(\theta_j-\theta_i)\\ &\quad +\frac{K_2}{N}\sum_{j\notin N_i}S_{ij}\sin(\theta_j-\theta_i),\end{aligned} \quad (7)$$

where $K_1,K_2>0$ in this model. $S_{ij}>0$ (i.e., the j-th node is the neighbor of the i-th node) would make the connected nodes cluster into the same community. $S_{ij}<0$ (i.e., the j-th node is not the neighbor of the i-th node) would make the unconnected nodes evolve far away.

The algorithm flow (i.e., DE_M) is given as follows.

Step 1 Initializing. The initial phases (i.e., θ_i) are generated randomly from $[0,2\pi]$ and initial frequencies (i.e., ω_i) are generated randomly from $[-0.01,0.01]$. Give two values for the parameters K_1 and K_2.

Step 2 Dynamic evolving. The phases of the nodes evolve according to Eq. (7).

Step 3 Detecting. Let phases mod 2π obtain better results. Then, plot the phases of the nodes evolving with time and find the different communities.

Parameter selection. K_1 is the coefficient of the positive coupling, and K_2 is the coefficient of negative coupling. Fix the value of K_2, regulate the value of K_1 according to the phase–time figure to obtain satisfactory results in the following simulations. Note that the gap of the K_1 and K_2 cannot be too big, or the phases of the nodes would realize synchronization when the value of K_1 is too large.

2.2. Improved dynamic evolutionary algorithm based on modularity matrix

We still require DE_M to find suitable parameters K_1 and K_2 for efficient community detection. Thus, we improve it further in this subsection.

In Ref. [31], they have shown that further enhancement in synchronization can be achieved by exploiting the information contained in the overall topology, i.e., by scaling the coupling strength to the load of each link. Now, we can obtain the following network equation:

$$\dot{\theta}_i = \omega_i + \frac{\sigma}{\sum_{j\in N_i} l_{ij}^{\alpha}} \sum_{j\in N_i} l_{ij}^{\alpha} \sin(\theta_j - \theta_i), \tag{8}$$

where l_{ij} is the load of the link connecting nodes i and j in the chosen network, and α is a tunable parameter. It has been shown that the ability of a dynamical network, like that in Eq. (8), to maintain a synchronization state crucially depends on the value of the parameter α. However, the complexity of l_{ij} is $O(mN)$.

In this paper, we further improve the formulation as follows (denoted as DE_M$^{\alpha}$):

$$\begin{aligned}\dot{\theta}_i = {} & \omega_i + \frac{K_1}{D} \sum_{j\in N_i} S_{ij}^{\alpha_1(t)} \sin(\theta_j - \theta_i) \\ & + \frac{K_2}{D} \sum_{j\notin N_i} |S_{ij}|^{\alpha_2(t)} \sin(\theta_j - \theta_i).\end{aligned} \tag{9}$$

Here, $D = \sum_{j\in N_i} S_{ij}^{\alpha_1(t)} + \sum_{j\notin N_i} |S_{ij}|^{\alpha_2(t)}$. The complexity of S_{ij} is only $O(m+N)$. Besides, the symbols of the S_{ij} divide the nodes into two parts. One part is a neighbor set and controlled by the coefficients K_1 and $\alpha_1(t)$, and the other part is the non-neighbor set and controlled by K_2 and $\alpha_2(t)$. Different parameters can obtain rich dynamic behaviors. $\alpha_1(t)$ and $\alpha_2(t)$ are time-dependent exponents, such that $\alpha_1(0) = 0$ and $\alpha_2(0) = 0$. $\alpha_1(t)$ and $\alpha_2(t)$ are stepwise increasing functions of time, which are defined as

$$\alpha_1(t) = \lfloor \frac{t}{T} \rfloor \delta_1, \tag{10}$$

$$\alpha_2(t) = \lfloor \frac{t}{T} \rfloor \delta_2. \tag{11}$$

Here, δ_1 and δ_2 are the step size, and T is a time step. At $t = 0$, we obtain $\alpha_1(0) = 0$ and $\alpha_2(0) = 0$, which means that $S_{ij}^{\alpha_1(t)}$ and $|S_{ij}|^{\alpha_2(t)}$ are equal to 1. Thus, $D = N$. Equation (9) would be reduced to the following form:

$$\dot{\theta}_i = \omega_i + \frac{K_1}{N} \sum_{j\in N_i} \sin(\theta_j - \theta_i) + \frac{K_2}{N} \sum_{j\notin N_i} \sin(\theta_j - \theta_i). \tag{12}$$

It is the same as the model in Ref. [28], which means equation (9) is the generalization of the model in Ref. [28] and equation (9) can gain rich dynamic behaviors.

The algorithm flow (i.e., DE_M$^{\alpha}$) is given as follows.

Step 1 Initializing. The initial phases (i.e., θ_i) are generated randomly from $[0, 2\pi]$ and initial frequencies (i.e., ω_i) are random from $[-0.01, 0.01]$. Give values for the parameters K_1, K_2, T, δ_1, and δ_2.

Step 2 Dynamic evolving. The phases of the nodes evolve according to Eqs. (9)–(11).

Step 3 Detecting. Let phases mod 2π to obtain better results. Then, plot the phases of the nodes evolving with time and find the different communities.

Parameter selection. Equation (12) can be further rewritten as

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^{N} K_{ij} \sin(\theta_j - \theta_i), \tag{13}$$

where

$$K_{ij} = \begin{cases} \dfrac{K_1}{N}, & j \in N_i, \\ \dfrac{K_2}{N}, & j \notin N_i. \end{cases} \tag{14}$$

In Ref. [32], they have proved that the model (Eq. (12)) can realize synchronization as K_1 and K_2 are large enough. Thus, in the simulations, we set K_1 and K_2 to be large enough to ensure that equation (9) can realize synchronization when $\alpha(0) = 0$. Here, δ_1 is the coefficient of the neighbor set and δ_2 is the coefficient of the non-neighbor set, and δ_1 is larger than δ_2 in the simulations to obtain satisfactory results. In the simulations, we select δ_1 from interval [0.05, 10] and δ_2 from [0.01, 30] for different networks.

3. Performance evaluations

Test on Zachary's karate club network The Karate club network is a typical network literature example used as a benchmark for community detection.[33] This is a network with 34 karate club members and 78 connections between members. The network is separated into two parts because a dispute arose between the club's administrator and its karate teacher. From Fig. 1(a), we can see that the phases of different nodes are clustered into two parts according to DE_M. The corresponding communities are detected with different colors in Fig. 1(b).

Figures 1(c)–1(f) show that 34 nodes will split into two communities according to DE_M$^{\alpha}$. We can see that the phase will synchronize firstly and be stable in two communities evolving with time and different α_1 and α_2. Synchronization in the same community (Fig. 1(c)) is enhanced and different communities are more obvious than DE_M in Fig. 1(a).

In Figs. 1(a) and 1(c), the phases of the nodes can cluster more closely when they are stable compared to Refs. [23] and [28]. DE_M and DE_M$^{\alpha}$ do not need to delete the leaf nodes and calculate the membership degree of several nodes in Ref. [28].

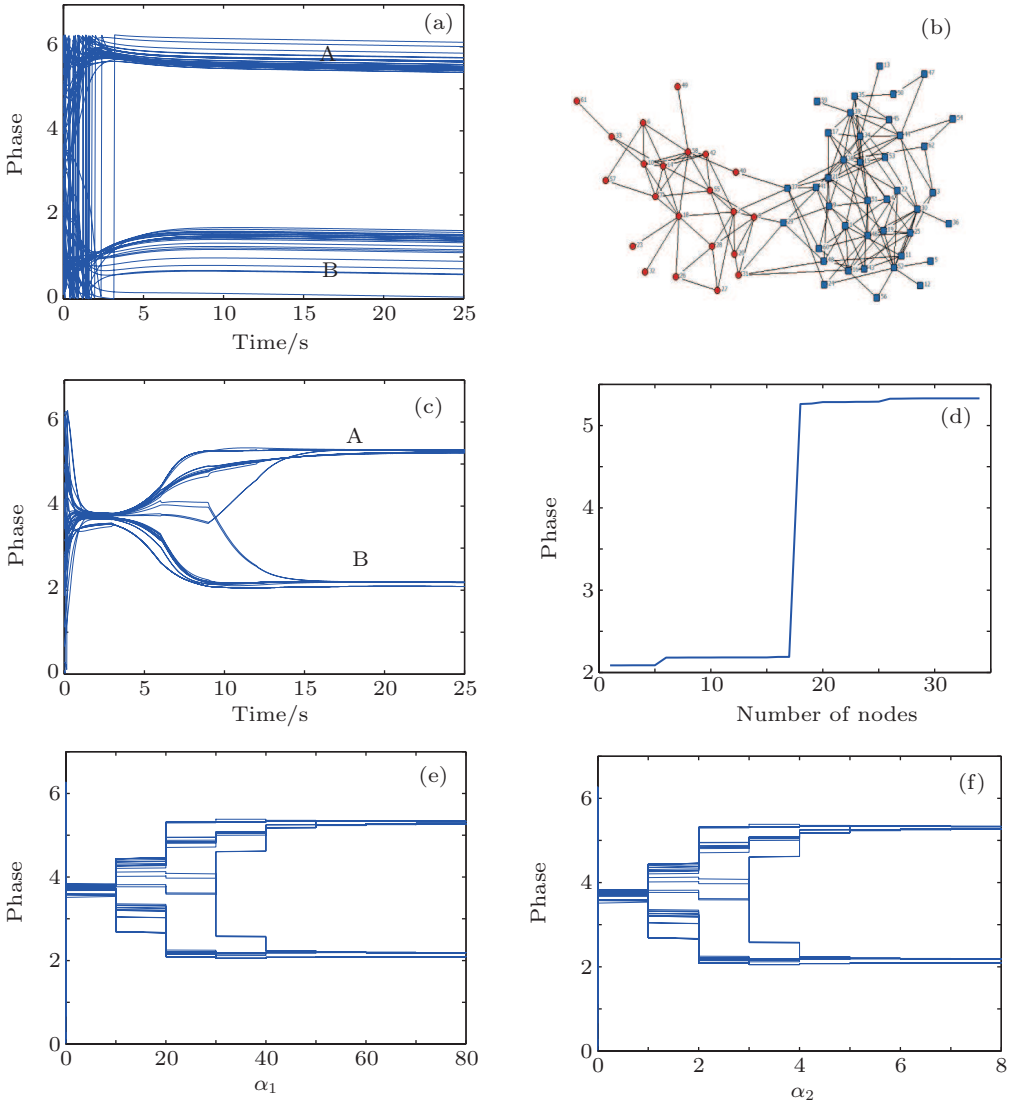


Fig. 1. (color online) Zachary's karate club network. (a) Phases of the nodes evolving with time based on DE_M. $K_1 = 45$ and $K_2 = 65$. (b) The colors correspond to the partitions given by DE_M. The individuals are partitioned into two parts denoted by a circle and a square, respectively. (c) Phases of the nodes evolving with time based on DE_M$^{\alpha}$. $K_1 = 80$, $K_2 = 1$, $T = 3$, $\delta_1 = 10$, and $\delta_2 = 1$. (d) Phases of the nodes will split into two stages based on DE_M$^{\alpha}$. (e) and (f) Corresponding phases of the nodes evolving with different α_1 and α_2 based on DE_M$^{\alpha}$, respectively.

Test on dolphins network Here, we turn to the community network of dolphins, which was constructed from observations of a community of 62 bottlenose dolphins over a period of seven years (from 1994 to 2001).[34] The phases of nodes in the dolphin network are clustered into two communities and the detection results are shown in Fig. 2.

Similar to Ref. [28], we obtain the histogram plot. The phases of the nodes can be divided into h (e.g., $h = 63$) subintervals with length $2\pi/h$. The number of nodes in each subinterval can be obtained (Fig. 2(d)). The two peaks in the figure depict the two communities.

Test on Journal network The journal index network is constructed by 40 journals from four different fields: physics, chemistry, biology, and ecology. 189 edges connect nodes if at least one article from a journal cites an article in the other journal during 2004.[35] The journal index network is ultimately clustered into four communities and the results are shown in Fig. 3.

Test on GN benchmark The famous benchmark of community detection is a class of networks introduced by Girvan and Newman (GN).[10] The network has 128 nodes, which are divided into four groups, each with 32 nodes. The detection results based on DE_M and DE_M$^{\alpha}$ with different parameters are shown in Fig. 4. The four communities are detected clearly by our proposed algorithms when $\mu = 0.4$ (μ is the mixing parameter, which expresses the ratio between the external degree of a node with respect to its community and the total degree of the node). Each node has a probability p_{in} of being connected to nodes of its group and a probability p_{out} of being connected to nodes of different groups.[36]

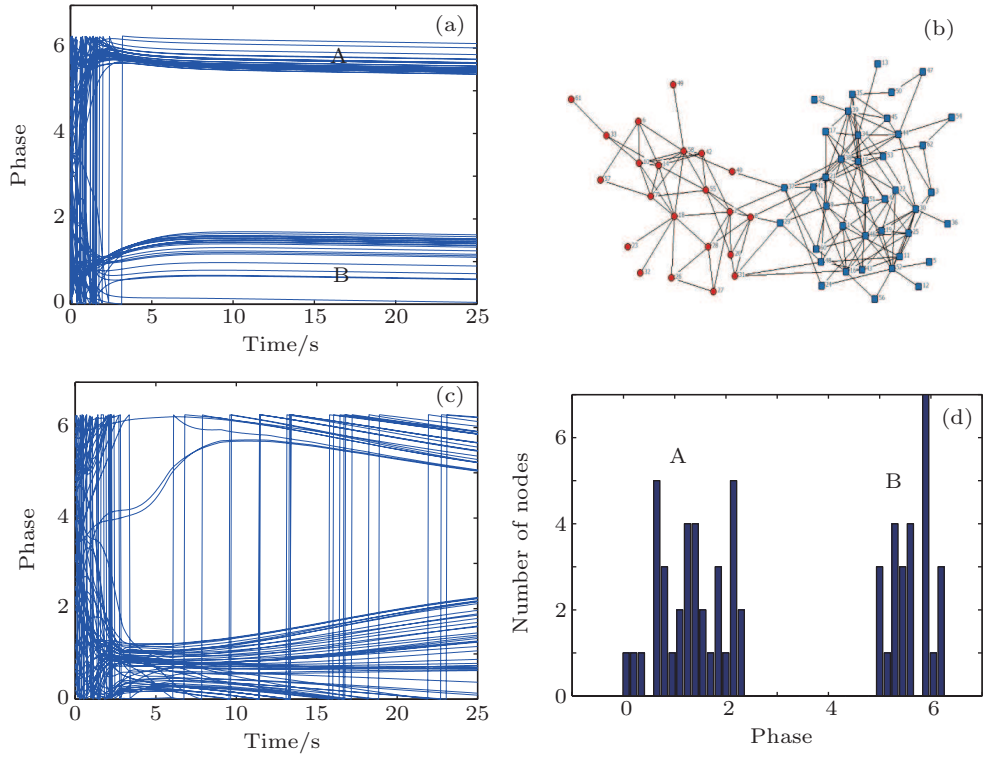


Fig. 2. (color online) Dolphin network. (a) Phases of the nodes evolving with time based on DE_M. $K_1 = 140$ and $K_2 = 10$. (b) The colors correspond to the partitions given by DE_M. The individuals are partitioned into two parts denoted by a circle and a square, respectively. (c) Phases of the nodes evolving with time based on DE_M$^{\alpha}$, here $K_1 = 60$, $K_2 = 0.1$, $T = 3$, $\delta_1 = 10$, $\delta_2 = 0.01$. (d) Distribution of the phases of nodes in $[0, 2\pi)$ based on DE_M$^{\alpha}$ by histogram plot.

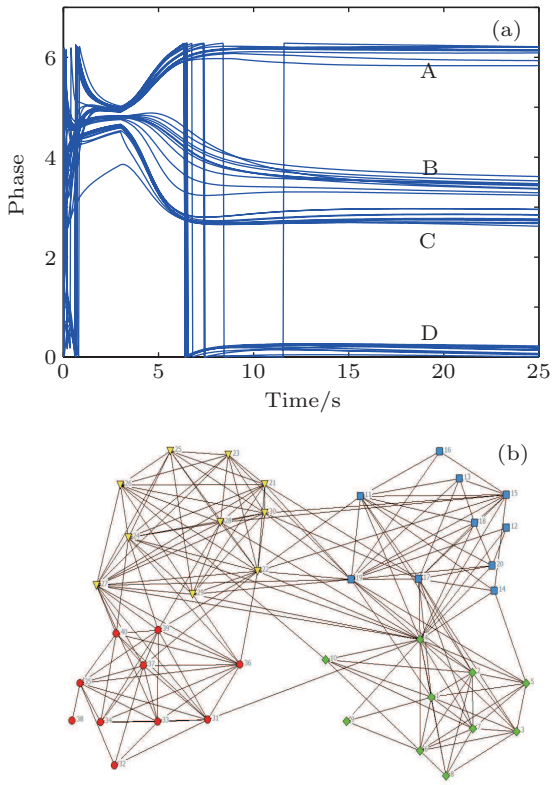


Fig. 3. (color online) Journal network. (a) Phases of the nodes evolving with time based on DE_M$^{\alpha}$. $K_1 = 65$, $K_2 = 1$, $T = 3$, $\delta_1 = 10$, and $\delta_2 = 0.05$. (b) Colors correspond to the partitions given by DE_M$^{\alpha}$. The individuals are partitioned into four parts denoted by a diamond, square, downtriangle, and circle, respectively.

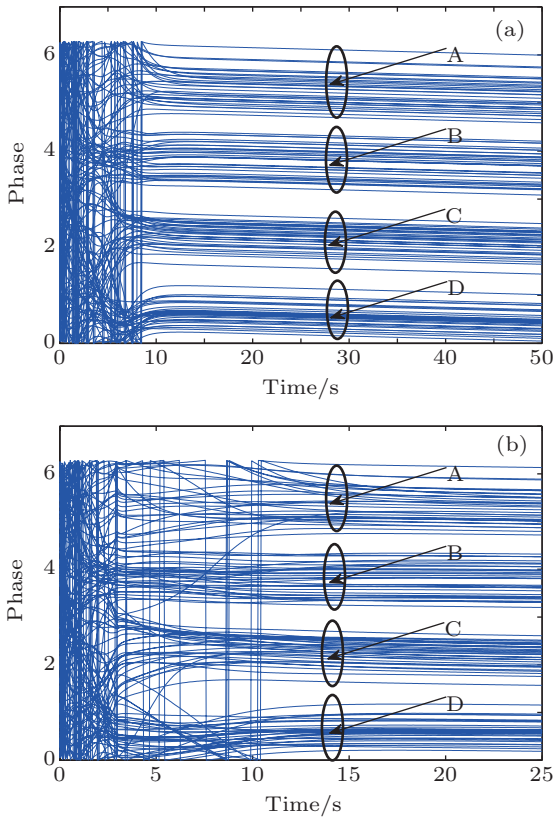


Fig. 4. (color online) GN benchmark with $\mu = 0.4$, $p_{\text{in}} = 0.3099$, and $p_{\text{out}} = 0.0665$. (a) Phases of the nodes evolving with time based on DE_M. $K_1 = 45$ and $K_2 = 65$. (b) Phases of the nodes evolving with time based on DE_M$^{\alpha}$. $K_1 = 70$, $K_2 = 15$, $T = 3$, $\delta_1 = 30$, and $\delta_2 = 0.1$.

4. Conclusions

The synchronization of the Kuramoto oscillators is usually adopted to reveal the community structure. Different from the previous work about the relationship between the community detection and the network dynamic clustering algorithms, in this paper, the coupled Kuramoto model is generalized based on the modularity matrix and the community of the network can realize synchronization or semi-synchronization. The main contributions are: (i) they do not need to delete the leaf nodes and further decide which community the special nodes belong to after phase dynamic synchronization; (ii) the time complexity of DE_M$^{\alpha}$ $(O(m+N))$ is much lower compared to the edge betweenness $(O(mN))$; (iii) the phases of the nodes can split into different communities according to the improved Kuramoto model; and (iv) the parameters can be easily controlled due to the adjustable α. In the future, we hope to find more efficient algorithms which can cluster the nodes that are more concentrated in the same community and more sparse between different communities.

References

[1] Buchanan M 2002 Nexus (New York-London: W. W. Norton)
[2] Chen J, Jiao L, Wu J and Wang X 2010 *Nonlinear Analysis: Real World Applications* **4** 3045
[3] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 *Physics Reports* **4** 175
[4] Strpgatz S H 2001 *Nature* **6825** 268
[5] Li H J, Wang Y, Wu L Y, Zhang J and Zhang X S 2012 *Phys. Rev. E* **1** 016109
[6] Wang L, Wang J, Shen H W and Cheng X Q 2013 *Chin. Phys. B* **22** 108903
[7] Sun F L and Zhu W 2013 *Chin. Phys. B* **22** 110204
[8] Yan X and Wu Y 2012 *Chin. Phys. Lett.* **29** 128901
[9] Gong M, Ma L, Zhang Q and Jiao L 2012 *Physica A* **391** 4050
[10] Girvan M and Newman M 2002 *Proc. Natl. Acad. Sci. USA* **12** 7821
[11] Radicchi F, Castellano C, Cecconi F, Loreto V and Parisi D 2004 *Proc. Natl. Acad. Sci. USA* **101** 2658
[12] Zhou H 2003 *Phys. Rev. E* **6** 061901
[13] Pan Y, Li D H, Liu J G and Liang J Z 2010 *Physica A* **14** 2849
[14] Capocci A, Servedio V D, Caldarelli G and Colaiori F 2005 *Physica A* **2** 669
[15] Li H J, Zhang J, Liu Z P, Chen L and Zhang X S 2012 *Europhys. Lett.* **6** 1
[16] Reichardt J and Bornholdt S 2004 *Phys. Rev. Lett.* **21** 218701
[17] Wu F and Huberman B A 2004 *Eur. Phys. J. B* **2** 331
[18] Donetti L and Munoz M A 2004 *J. Stat. Mech.: Theor. Exp.* **10** P10012
[19] Duch J and Arenas A 2005 *Phys. Rev. E* **2** 027104
[20] Wu J, Wang X and Jiao L 2012 *Physica A* **391** 508
[21] Boccaletti S, Ivanchenko M, Latora V, Pluchino A and Rapisarda A 2007 *Phys. Rev. E* **75** 045102
[22] Wu J, Li L, Wang X, Yu X and Jiao L 2013 *Chin. Phys. B* **22** 060507
[23] Wu J, Lu R, Jiao L, Liu F, Yu X, Wang D and Sun B 2013 *Physica A* **392** 1287
[24] Kuramoto Y 1984 *Chemical Oscillations, Waves and Turbulence* (Berlin: Springer-Verlag)
[25] Pluchino A, Latora V and Rapisarda A 2005 *Int. J. Mod. Phys. C* **4** 515
[26] Khadivi A, Rad A A and Hasler M 2010 *Proceedings of* 2010 *IEEE International Symposium on Circuits and Systems*, May 30–June 2, 2010 Paris, France, pp. 3777–3780
[27] Moujahid A, D'Anjou A and Cases B 2012 *Chaos* **9** 1171
[28] Wu J, Jiao L, Jin C, Liu F, Gong M, Shang R and Chen W 2012 *Phys. Rev. E* **85** 016115
[29] Newman M E J and Girvan M 2004 *Phys. Rev. E* **2** 026113
[30] Newman M E J 2006 *Proceedings of the National Academy of Sciences of the United States of America* **23** 8577
[31] Chavez M, Hwang D U, Amann A, Hentschel H G E and Boccaletti S 2005 *Phys. Rev. Lett.* **94** 218701
[32] Ha S Y, Ha T and Kim J H 2010 *Physica D* **17** 1692
[33] Zachary W W 1977 *Journal of Anthropological Research* **33** 452
[34] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E and Dawson S M 2003 *Behavior Ecol. Sociobiol.* **4** 396
[35] Zhang D, Xie F, Zhang Y, Dong F and Hirota K 2010 *Physica A* **389** 5319
[36] Lancichinetti A and Fortunato S 2009 *Phys. Rev. E* **80** 056117