Comparative definition of community and corresponding identifying algorithm

Yanqing Hu, Hongbin Chen, Peng Zhang, Menghui Li, Zengru Di, and Ying Fan[*]
*Department of Systems Science, School of Management, Center for Complexity Research, Beijing Normal University,
Beijing 100875, China*

A comparative definition for community in networks is proposed, and the corresponding detecting algorithm is given. A community is defined as a set of nodes, which satisfies the requirement that each node's degree inside the community should not be smaller than the node's degree toward any other community. In the algorithm, the attractive force of a community to a node is defined as the connections between them. Then employing an attractive-force-based self-organizing process, without any extra parameter, the best communities can be detected. Several artificial and real-world networks, including the Zachary karate club, college football, and large scientific collaboration networks, are analyzed. The algorithm works well in detecting communities, and it also gives a nice description of network division and group formation.

I. INTRODUCTION

Many physicists have become interested in the study of networks describing the topologies of a wide variety of systems [1–3], such as the World Wide Web [4], social and communication networks [5,6], biochemical networks [7], and many more. Many networks are found to divide naturally into communities. Nodes belonging to a tight-knit community are more than likely to have other properties in common. In the world wide web, community analysis has uncovered thematic clusters. In biochemical or neural networks, communities may be functional groups. As a result, the problem of identification of communities has been the focus of many recent efforts. Many different algorithms are proposed [8–23] (see [24] for a review).

Communities within networks can loosely be defined as subsets of nodes which are more densely linked, when compared to the rest of the network. Modularity Q [25] was presented as an index of community structure and now has been widely accepted [8,9,14,22] as a measure for the communities. Modularity Q was introduced by Newman and Girvan as follows:

$$Q = \sum_r (e_{rr} - a_r^2), \tag{1}$$

where e_{rr} is the fraction of links that connect two nodes inside the community r, a_r is the fraction of links that have one or both vertices inside the community r, and the sum extends to all communities r in a given network. Note that this index provides a quantitative measurement to decide the best division of networks. The larger the value of Q, the more accurate is a partition into communities. So maximizing modularity Q can also detect communities. Actually, there are already many algorithms of maximizing Q such as extremal optimization (EO) [22], Greedy algorithm [12], and other optimal algorithms. There are also many other algorithms to identify communities in complex networks such as the Newman-Girvan (GN) algorithm [17,25], random walk

method [10], edge clustering coefficient method [20], and spectral analysis [8]. When the methods can only produce the dendrogram of the community structure, the best partition is usually obtained by maximizing modularity Q. Unfortunately, the modularity Q maximization problem was proved to be an NPC problem [26]. Moreover, it has been proved that modularity Q measurement may fail to identify modules smaller than a scale which depends on the total number L of links of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined [27].

There are also other community definitions based on the topology of networks, such as self-referring definitions and comparative definitions. The basic self-referring definition is a clique, defined as a subgroup of a graph containing more than two nodes where all the nodes are connected to each other by means of links in both directions. In other words, this is a fully connected subgraph. This is a particularly strong definition and rarely fulfilled in real sparse networks for larger groups [28]. The other referring community definition is the k core, which is defined as a subgraph in which each node is adjacent to at least a minimum number k of the other nodes in the subgraph. It is weaker than a clique, but it is very hard to find the optimal k when we want to detect the best partition of the network. Comparative definitions are given on the basis of links comparison. There are three kinds of comparative definitions which are called the *LS*-set, strong, and weak community definitions. The *LS* set is defined as a set of nodes in which each of its subsets has more ties to its components within the set than outside [29]. The *LS*-set definition is also quite stringent. Moreover, it is a very tough problem to detect all the *LS* sets in a network. In order to relax the constraints, Raddichi *et al.* [20] proposed the strong and weak definitions. In a strong community, each node has more connections within the community than with the rest of the network, and in a weak community the sum of all degrees within the community is larger than the sum of all degrees toward the rest of the network. Based on these comparative definitions, the self-contained algorithm is developed, which is similar to the GN algorithm for finding strong or weak communities in a network.

---

*Author for correspondence: yfan@bnu.edu.cn

In this article, following the basic idea of comparative definitions, we define community as follows: a community is a set of nodes, and each node's degree inside the community should be larger than or at least equal to its degree link to any other community. This definition is different from other comparative definitions. The strong, weak, and *LS*-set definitions are presented by comparing the degree in the community with the degree in the whole rest network. But our community definition is designed by comparing the degree in the community with the degree in each rest community, not the whole rest network.

Then how do we detect the communities in a network based on our definition? Obviously, whether a node belongs to a community or not is determined by its connections. We can define the attractive force of a community to a node by the links connect them. Employing an attractive-force-based self-organize process, the community structures can be detected without any extra parameter. The algorithm also gives a nice description of the affection of a community to a node and group formation process.

This paper is organized as follows. Section II gives our comparative definition for communities in networks. Then in Sec. III, the corresponding algorithm is given in details. The application of the definition and the algorithm in *ad hoc*, Zachary karate club, college football, and large scientific collaboration networks are presented in Sec. IV. Some concluding remarks are given in Sec. V.

II. QUANTITATIVE DEFINITIONS OF COMMUNITY

A. Previous comparative definition

The most important comparative definitions of community are strong and weak definitions, which are proposed by Raddichi *et al.* [20]. Suppose there is a network G which has n nodes and it can be represented mathematically by an adjacency matrix A with elements $A_{i,j}=1$ if there is an edge from i to j and $A_{i,j}=0$ otherwise.

Definition of community in a strong sense. The subnetwork V is a community in a strong sense if for any i belonged to V we have

$$\sum_{j \in V} A_{i,j} > \sum_{j \in (G-V)} A_{i,j}. \tag{2}$$

Definition of community in a weak sense. The subnetwork V is a community in a weak sense if we have

$$\sum_{i,j \in V} A_{i,j} > \sum_{i, \in V, j \in (G-V)} A_{i,j}. \tag{3}$$

Obviously, the strong community definition concerns the situation of every node, but the weak sense takes a community as a whole. From the strong (weak) definition of community we can easily get that if $V_1, V_2 \subseteq G$ satisfy the strong (weak) definition, then we have that $V_1 \cup V_2$ also satisfy the strong (weak) definition. Raddichi *et al.* [20] call this phenomena as self-contained and use the self-contained algorithm to detect communities, which is similar to the GN algorithm for finding strong or weak communities in a network. Instead of cutting the edge with largest link be-

tweenness, the self-contained algorithm cuts the edge with the smallest value of the edge-clustering coefficient. They introduced an approach to calculate the edge-clustering coefficient based on circles of order g ($g=3$, triangles, and $g=4$, squares) in the networks. In this paper, we mainly discuss the self-contained algorithm with edge-clustering coefficient of order 3 [self-contained algorithm (g3) for short].

B. New community definition

Inspired by the above strong and weak definitions, we define the community as follows.

Definition of Community. If $V_1, V_2, \ldots, V_m$ are m communities of G, V_k, $k=1,2,\ldots,m$, should satisfy that

$$\bigcup_{k=1}^{k=m} V_k = G \tag{4}$$

and

$$\forall j \in V_k, \quad \sum_{i \in V_k} A_{i,j} \geq \max\left\{\sum_{i \in V_t} A_{i,j}, t=1,2,\ldots,m\right\}. \tag{5}$$

This definition can be summarized as follows: a community should satisfy that each node's degree inside the community should not be smaller than the node's degree toward any other community. The same as the strong sense, our definition also focuses on the situation of the node. But different from comparing the degree in the community with the degree in the whole rest network, our definition compares the degree in the community with the degree in each rest community instead of the whole rest network. Obviously, our definition is weaker than the strong definition. Here we can also give another most weak community definition: in a community, the sum of all degrees inside the community should not be smaller than the sum of degrees toward any one other community. The same as the weak sense, our most weak community definition focuses on the case of community instead of the single node. The difference between the weak definition and our most weak definition is that the weak definition compares the sum of degrees inside the community with the sum of degrees towards the whole rest network, but the most weak one compares the sum of degrees inside the community with the sum of degrees towards any other community. In the following discussion, we only deal with the new definition given by formula (5) and mainly compare it with the strong definition.

Let us first give some analytical and quantitative comparisons of different definitions of community on artificial networks following the same lines of calculations in [30] for strong and weak definitions. Each network has $n=128$ nodes divided into 4 communities of 32 nodes each. Edges between two nodes are introduced with different probabilities depending on whether the two nodes belong to the same community or not: every node has $\langle k_{intra} \rangle$ links on average to its fellows in the same community, and $\langle k_{inter} \rangle$ links to the outer world, keeping $\langle k_{intra} \rangle + \langle k_{inter} \rangle = 16$. We also can say each pair of nodes in the same community has an edge with probability $p_{in} = \frac{k_{intra}}{(N_{in}-1)}$ and each pair of nodes in the different community has an edge with probability $p_{out} = \frac{k_{inter}}{N_{out}}$, where $N_{in}=32$ and
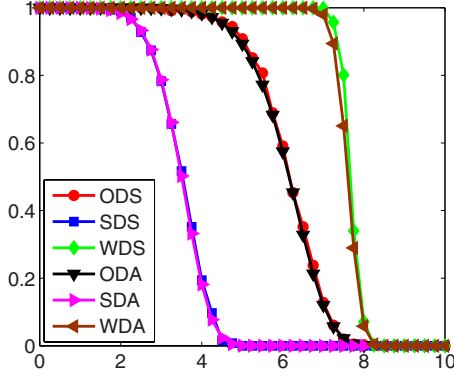
FIG. 1. (Color online) Proportion of networks that all four presumed networks satisfy the corresponding definitions with the changes of out-degree. Each point denotes the average value of 1000 realizations. In the plot, ODS, SDS, WDS, ODA, SDA, and WDA mean our definition simulation, strong definition simulation, weak definition simulation, our definition analysis, strong definition analysis, and weak definition analysis, respectively.

$N_{out} = 96$. Then the probability that node i has u inward edges is given by

$$p(u, N_{in}) = C_{N_{in}-1}^u p_{in}^u (1 - p_{in})^{N_{in}-1-u} \qquad (6)$$

and the probability that node i has v outward connections is given by

$$q(v, N_{out}) = C_{N_{out}}^v p_{out}^v (1 - p_{out})^{N_{out}-v}, \qquad (7)$$

where C in the above two equations is the binomial coefficient. Therefore, the probability that node i has u inward edges and satisfies our definition is given by

$$P(u) = p(u, N_{in}) \left(\sum_{v_1=0}^u \sum_{v_2=0}^u \sum_{v_3=0}^u q(v_1, N_{in}) q(v_2, N_{in}) q(v_3, N_{in}) \right). \qquad (8)$$

Hence, the probability that all the presumed communities satisfy the our definitions is given by

$$P = \left(\sum_{u=0}^{N_{in}-1} P(u) \right)^n. \qquad (9)$$

We have shown the change of percentage that all four presumed communities satisfy the corresponding definitions (as shown in Fig. 1). We could find that our definition is in the middle of the strong and weak definitions given by Raddichi *et al.* [20].

III. ALGORITHM

In order to detect the community structure under our new definition, we set each node and its random half of neighbors to be a community initially. Then we define the attractive force by the connections among nodes and let the communities be self-organized with the forces. When the community structure becomes fixed, the survivors will be the best partitions which satisfy the above definition naturally.

Let $F_{k,i}$ denotes the attractive force of community k to node i and $F_{k,i}$ can be calculated by the formula

$$F_{k,i} = \sum_{j \in V_k} A_{i,j}. \qquad (10)$$

Then our algorithm is defined as follows.

(1) We initially set each node and its random half neighbors to be a community. If a node has h neighbors and h is odd, we let the node and its random $\frac{h+1}{2}$ nodes as a community. If two or more than two communities are the same, just keep one of them. So after the first step the network is partitioned to n or less than n overlapping communities. n is the number of nodes in the network.

(2) Calculate $F_{k,i}$ for all k and i.

(3) For every node, move it into the community or communities with the largest attractive force, respectively, at the same time.

(4) Check all communities; if two or more than two communities are the same, just keep one of them.

(5) Repeat steps 2–4 until sufficient N steps or the partition be fixed.

The time complexity of our algorithm is $O(n^2)$. Step 1 runs in time $O(dn)$, step 2 in $O(n^2)$, step 3 in $O(n^2)$, step 4 in $O(n^2)$, and the repeated time in step 5 is uncertain, where d is the average degree. According to the numerical experiments in artificial networks, around ten repeating steps, the partition will be fixed. So we think the time complexity is $O(n^2)$.

Even our definition of communities is not a self-contained one such as strong and weak definitions; there should be more than one partition that may satisfy our community definition. So we keep some stochastic factors in our initial partition and run the algorithm several times. Then we could report the average result or choose the best one from all the partitions. Here we introduce another indicator for evaluating the partitions. We think the best partition should satisfy that there are more connections inside the communities and less connections outside the communities. So we use the proportion of average connection density inside the communities and the connection density outside the communities to measure how reasonable a partition is. This kind of measurement can be defined as follows. Suppose the network contains n nodes and L connections and is partitioned to m communities. n_i, $i=1,2,\ldots,m$, denotes the number of nodes in the ith community and L_i, $i=1,2,\ldots,m$, denotes the number of connections in the ith community. Then the average connection density inside the communities is

$$D_{in} = \frac{2\sum_{i=1}^m L_i}{\sum_{i=1}^m n_i^2 - n} \qquad (11)$$

and the connection density outside the communities

$$D_{out} = \frac{2L - 2\sum_{i=1}^m L_i}{n^2 - \sum_{i=1}^m n_i^2}. \qquad (12)$$

Then the measurement H can be defined as $H = \frac{D_{in}}{D_{out}}$ and when there only one community $H=0$. Obviously, larger H means a more reasonable partition.
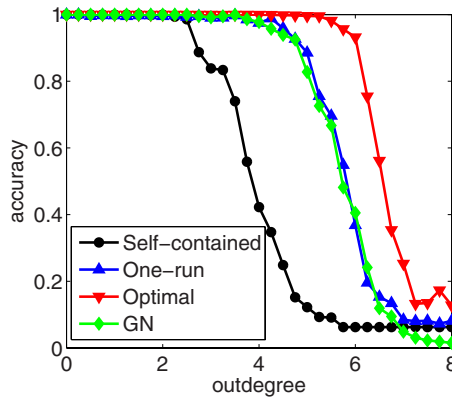
FIG. 2. (Color online) The accuracies of our algorithms, GN algorithm, and self-contained algorithm (g3). From the plot we can see that the accuracy of the one-run algorithm is similar to the GN algorithm. The partition of the optimal algorithm with the aid of measurement H is better than GN algorithm when the out-degree becomes larger, where we run 15 times for each network for optimum. Each point is the average of 20 realizations of networks.

IV. APPLICATION IN *ad hoc* AND REAL NETWORKS

A. Algorithm on artificial networks

In order to test our algorithm, we apply it to artificial networks with 128 nodes and 4 predetermined communities, which are the same kind networks in Sec. II. The accuracy of the algorithm is evaluated by the similarity function S [31]. For each given out degree $\langle k_{inter} \rangle$, we produce 20 realizations of networks. Then, for each network, we first run the algorithm one time and give the average accuracy of 20 networks as one run shown in Fig. 2. Then we run the algorithm 15 times for each network and choose the best partition with the aid of indicator H. The average accuracy of 20 networks is also shown as optimal in Fig. 2. Comparing our algorithm with GN and self-contained algorithms [17,20,25], we could find that the accuracy of the one-run algorithm is similar to the GN-algorithm and better than the self-contained algorithm (g3), and the optimal algorithm is better than the GN, self-contained (g3), and one-run algorithms.

We also test, with the process of our algorithm, to what extent the partition satisfies our definition (see Fig. 3). For a given partition V_i, $i = 1, 2, \ldots, m$, we define its community degree (CD) as the ratio

$$ \text{CD} = \frac{\sum_{i=1}^{m} |\widetilde{V}_i|}{\sum_{i=1}^{m} |V_i|}, \tag{13} $$

where $\widetilde{V}_i$ denotes the subset of V_i, in which each node satisfies the requirement of our definition for community; that is, the node's interdegree is larger or equal to its intradegree between any other community. The numerical experiments results tell us when the community structure is not very fuzzy; the algorithm will finally produce a partition that satisfies our definition very well. The community degree tends to 1. When the community structure is very fuzzy, it is hard to find the partition that satisfies the definition exactly.
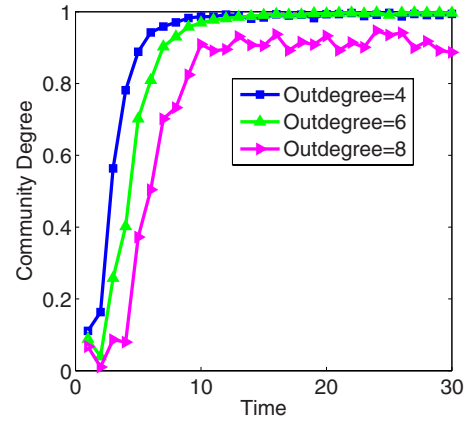

FIG. 3. (Color online) The evolution of community degree with the process of the algorithm. The results are for the one-run algorithm. We can see that when the community structure is not very fuzzy, the one-run algorithm services our community definition very well.

Recently Fortunato and Barthelemy [27] proved that modularity Q may fail to identify small communities and give a kind of network as shown in Fig. 4. We test our algorithm on this kind of networks. When each circle contains a clique with three or more than three nodes, our algorithm can detect all the predeterminate communities (circles) always.

Furthermore, we also have done a quantitative comparison of different definitions of community on a small graph with four complete subgraphs (as shown in Fig. 5). Without the red dashed link, four subgraphs are communities that fulfill the strong, weak, and our definitions and the H index is 9.0 (modularity $Q = 0.61$). But with the link, four subgraphs are no longer communities under the strong definition. With the red dashed link, the small graphs was divided into two communities $\{ACD, B\}$ or $\{ABD, C\}$, $H = 4.1$ (modularity $Q = 0.30$) by self-contained algorithm (g3), and four communities $\{A, B, C, D\}$, $H = 7.2$ (modularity $Q = 0.58$) by our algorithm.
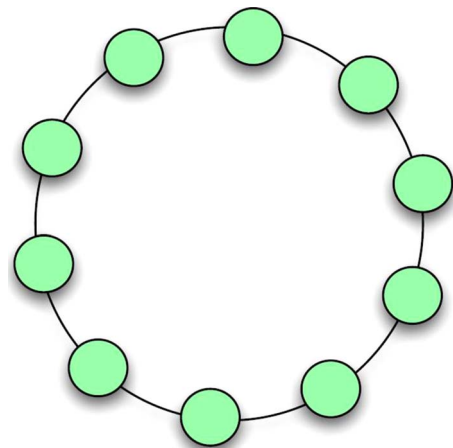

FIG. 4. (Color online) The circles represent the communities in which each pair of nodes are connected. The circles be connected to each other by the minimal number of links. The plot is cited from [27].
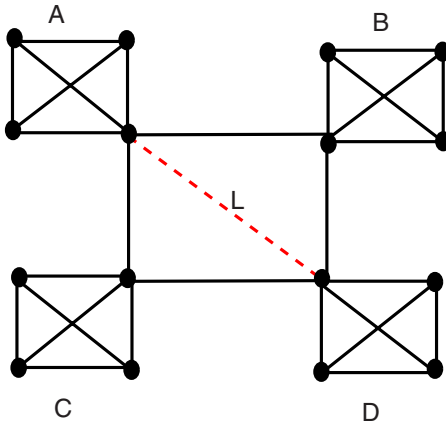
FIG. 5. (Color online) Small graph with four complete subgraphs. Without link L, four subgraphs are distinct communities that fulfill all the definitions. When link L is added, four subgraphs do not satisfy the strong definition, but they are communities under our definition.

B. Zachary karate club network

When apply our algorithm to a real network, first we use the popular Zachary karate club network [32], which is considered as a simple workbench for community-finding methodologies [17–19,21,25,33]. This network was constructed with the data collected observing 34 members of a karate club over a period of 2 years and considering friendship between members. By our algorithm, 3 communities are detected (as shown in Fig. 6). The partition is reasonable compared with the actual division of the club members.

As mentioned above, there may be many partitions that satisfy the requirement of our definition and the final partition is related to the initial conditions. For the karate club network, if we think the club division is caused by some leaders, such as leaders (nodes) 1, 33, and 34, and set the leaders and their random half neighbors as initial partition, then our algorithm will divide the network into two communities. That is consistent with the real division. If we set 1, 2, 33, and 34 as the leaders, our algorithm will also partition the
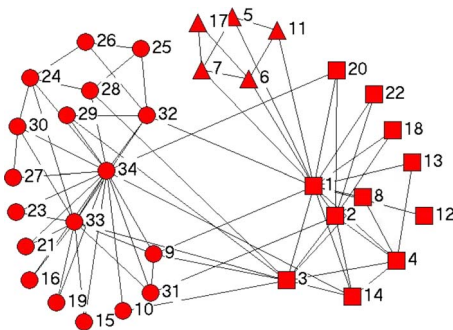

FIG. 6. (Color online) The community structure of the Zachary karate club network. Our algorithm detects three communities which are depicted by circles, squares, and triangles. When we set 1, 2, 33, and 34 as leaders, the partition is the same. But if we set 1, 33 and 34 as leaders, the network will be divided into two communities. Circles represent a community and the rest is another one, which corresponds to the actual division.

TABLE I. The accuracy of each detected community compared with the counterpart of real-world communities.

Conference name	Accuracy	GN accuracy
Atlantic Coast	1	0.9000
Big East	0.8000	1
Big10	1	1
Big12	1	0.9231
Conference USA	0.6429	0.9000
IA Independents	0	0
Mid American	0.8667	0.8667
Mountain West	1	0
Pac10	1	0.5556
SEC	1	0.7500
Sunbelt	0.4444	0.4444
Western Athletic	0.7273	0.7273
Accuracy	0.7901	0.6723

network into three communities, which is the same as the result without leaders. It is very interesting, with the process of group formation, that nodes 1,2 and nodes 33,34 combine and are in the same community, respectively. The other communities do not contain any nodes of 1, 2, 33, and 34. It implies that, if some leaders have contradictions and want to divide the network, some nodes will not always follow the leaders and may form other groups (see Fig. 6).

C. College football network

We also apply our algorithm to the collage football network which was provided by Newman [34]. The network is a representation of the schedule of Division I games for the 2000 season. Nodes in the network represent teams and edges represent regular-season games between the two teams they connect. What makes this network interesting is that it incorporates a known community structure. The teams are divided into 12 conferences [17]. Games are more frequent between members of the same conference than between members of different conferences. It is found that our overlapping algorithm identifies the conference structure with a high degree of success. We detect 12 communities in which 5 communities were detected exactly, the average accuracy is 0.79, and no node is overlapping. The GN algorithm associated with the Q function [25] gives the partition. It divides the football teams into 10 communities, and the average accuracy is 0.67. The results are shown in Table I.

D. Scientific collaboration network

Up to now we have only discussed examples of networks which are smaller than 200 nodes. It is necessary to test our algorithm on large networks. Newman composed a scientific collaboration network of scientists posting preprints on the condensed matter archive at www.arxiv.org form 1995 to 1999 [35] which contains 16 725 nodes and 47 594 edges (data download from [34]). We use the optimal algorithm with ($N=50$) and 100 runs. Then we got the division with
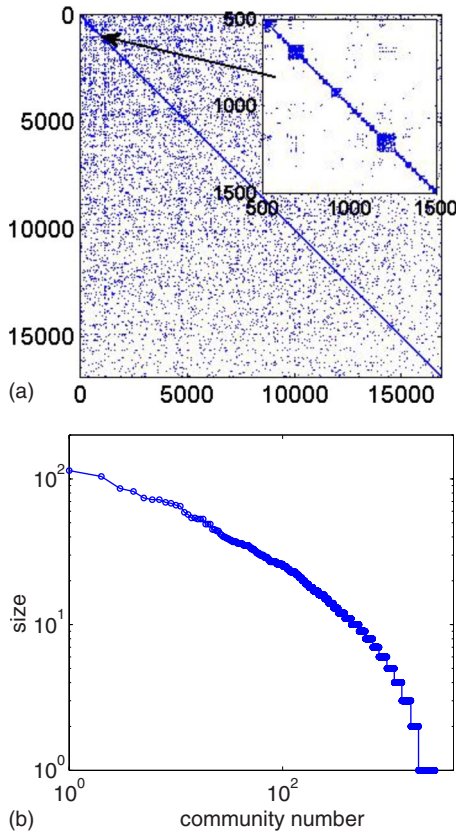
FIG. 7. (Color online) The reordered adjacent matrix (a) and Zipf plot of community size distribution (b) of a scientific collaboration network. The power-law density function for the upper tail is $p(s) \sim s^{-3.5}$.

maximal $H = 1754.8$ and 2676 communities. In the final division, 0.93 nodes satisfy our community definition and 158 nodes are overlapping. The community size distribution obeys a power law [36] with density function $p(s) \sim s^{-3.5}$ for the upper tail (see Fig. 7). In order to investigate the effect of repeating time N (see algorithm step 5), we recorded the community degree with the change of N. Here, community degree means the proportion of nodes that satisfy our definition. From Fig. 8, we can find that when $N > 10$ the community degree is near the maximum value, which is similar with the results on small artificial networks (see Fig. 3). It seems that the repeating time N to reach the final division of network is not dependent on the network size. It indicates that the complexity of our algorithm is $O(n^2)$.

V. CONCLUSION AND DISCUSSION

In this paper, we present a comparative community definition and the corresponding algorithm. A community should satisfy that each node's degree inside the community should be larger than or equal to the node's degree toward any other
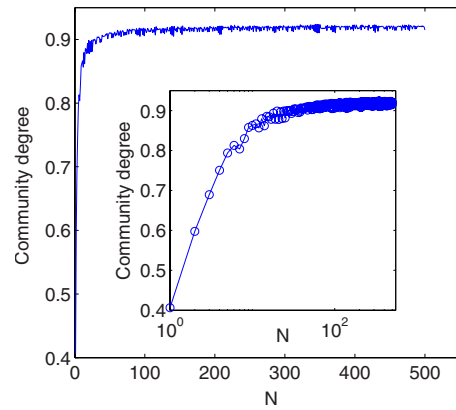

FIG. 8. (Color online) The evolution of the community degree of scientific collaboration network with the process of the one-run algorithm. We can see that for large N, the community degree cannot achieve 1 (maybe the evolution process enters a circle or chaos), but when $N = 10$, the community degree is near the maximum value 0.93.

community. Then we introduce the concept of attractive force and develop a self-organizing algorithm based on the comparing of attractive forces. The algorithm can detect the community structures without any extra parameter. In order to choose the best partition from several possible results, we also define an indicator H to evaluate the partitions. We apply the algorithm to artificial networks and some real-world networks such as the Zachary karate club network, college football network, and large scientific cooperation network. The algorithm work well in all networks. Furthermore, our community definition and identification algorithm can be generalized to weighted and directed networks easily.

Moreover, our algorithm can be use to predict network division when there are some contradictions between some leaders. In the algorithm, we can initially set some leaders and their random half neighbors to be the communities, respectively. Then the self-organizing process gives a nice description of leaders' affections. We think this partition technique has great potential for analyzing network structure.

In Sec. II, we give the most weak community definition: in a most weak sense, the sum of all degrees inside the community should not be smaller than the sum of degrees toward any other community. From the view of statistical physics, we think the most weak definition is also reasonable. Here we propose an open problem of finding a algorithm to detect the communities based on the most weak definition.

[1] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).

[2] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **401**, 130 (1999).

[5] S. Redner, Eur. Phys. J. B **4**, 131 (1998).

[6] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[7] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000).

[8] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[9] S. Lehmann and L. K. Hansen, e-print arXiv:physics/0701348.

[10] M. Latapy and P. Pons (unpublished).

[11] F. Wu and B. A. Huberman, Eur. Phys. J. B **38**, 331 (2004).

[12] A. Clauset, Phys. Rev. E **72**, 026132 (2005).

[13] S. Muff, F. Rao, and A. Caflisch, Phys. Rev. E **72**, 056107 (2005).

[14] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).

[15] C. P. Massen and J. P. K. Doye, Phys. Rev. E **71**, 046101 (2005).

[16] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori, Physica A **352**, 669 (2005).

[17] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2004).

[18] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).

[19] M. E. J. Newman and E. A. Leicht, Proc. Natl. Acad. Sci. U.S.A. **104**, 9564 (2007).

[20] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).

[21] J. P. Bagrow and E. M. Bollt, Phys. Rev. E **72**, 046108 (2005).

[22] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).

[23] J. Reichardt and S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004).

[24] L. Danon, J. Duch, A. Arenas, and A. Diaz-Guilera, J. Stat. Mech.: Theory Exp. (2005) P09008.

[25] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[26] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, e-print arXiv:physics/0608255 (2006).

[27] S. Fortunato and M. Barthelemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).

[28] G. Palla, I. Derenyi, I. Fakas, and T. Vicsek, Nature (London) **435**, 814 (2005).

[29] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, UK, 1994).

[30] C. Castellano, F. Cecconi, V. Loreto, and F. Radicchi, Eur. Phys. J. B **38**, 3111 (2004).

[31] Y. Fan, M. Li, P. Zhang, J. Wu, and Z. Di, Physica A **377**, 363 (2007).

[32] W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[33] L. Donetti and M. A. Munoz, J. Stat. Mech.: Theory Exp. (2004) P10012.

[34] http://www-personal.umich.edu/mejn/netdata/

[35] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).

[36] A. Clauset, C. R. Shalizi, and M. E. J. Newman, e-print arXiv:0706.1062.