# A Novel Algorithm Infomap-SA of Detecting Communities in Complex Networks

Fang Hu[1, 2] and Yuhua Liu[1]
[1]Schoolof Computer Science, Central China Normal University, Wuhan, 430079, China
[2]College of Information Engineering, Hubei University of Chinese Medicine, Wuhan 430065, China
Email: naomifang@mails.ccnu.edu.cn; yhliu@mail.ccnu.edu.cn

***Abstract*** —Community detection is one of the most important issues in complex networks. In this paper, integrating Infomap and Simulated Annealing (SA) algorithm, and based on the thought of optimizationof the modularity function, the authors are proposing a novel algorithm Infomap-SA for detecting community. In order to verify the accuracy and efficiency of this algorithm, the performance of this algorithm is tested on several representative real-world networks and a set of computer-generated networks by LFR-benchmark. The experimental results show that this algorithm can identify the communities accurately and efficiently, and has higher values of modularity and density and lower computable complexity than Infomap algorithm. Furthermore, the Infomap-SA is more suitable for community detection of large-scale network.

***Index Terms***—Community detection; modularity; density; infomap-simulated annealing algorithm; simulation test

## I. Introduction

In recent years, researches on community structure in complex networks have attracted great attentions from various fields, including biology, physics, sociology, mathematics, etc. [1]-[3]. In real world, many different kinds of networks such as natural, biological and social networks can be interpreted as complex networks [4], [5]. During 2002, Girvan and Newman found the property of community structure in complex network besides the small world and scale-free properties [6]. The identification of community structure in complex networks is one of the most important issues. Through researches on this problem, the unknown relationships between nodes can be revealed, and it has certain heuristic to explore the unknown world [7], [8].

There are many algorithms for detecting community at home and abroad at present. In 2004, Girvan and Newman presented a measure standard to evaluate the quality of community, which can be called the modularity $Q$ metric [9]. After that, a large number of algorithms were studied to detect community by optimizing the modularity [10], [11]. It is generally accepted that the value of the modularity is higher, the result of community detection is better. Newman established FN algorithm [12] based on local search and CNM algorithm [13]. Clauset presented a modularity maximization algorithm in practical contexts [14]. Lancichinetti studied a new modularity maximization algorithm in community detection [15].

Besides these algorithms based on modularity optimization, there are some other community detection algorithms. Jin *et al*. [16] presented an efficient detecting communities algorithm with self-adapted fuzzy C-means clustering. Gong *et al*. [17] proposed a novel heuristic density-based method for community detection. Crampes *et al.* [18] constructed a unified community detection, visualization and analysis method. Liu *et al*. [19] established a new clustering algorithm based on field in complex networks. Mu *et al*. [20] proposed a two-stage algorithm using influence coefficient for detecting the hierarchical, non-overlapping and overlapping community structure.

Based on minimum entropy, Rosvall and Bergstro presented the Infomap algorithm [21], which transforms the problem of community detection to information compression, and the problem of extreme value can be solved well. Infomap algorithm has the characteristic of running faster, and can detect large networks with no prior knowledge of the number of communities. However, there are some problems that the efficiency of clustering will decrease when the community structure is not obvious. Based on combinatorial optimization, simulated annealing (SA) algorithm was presented by Bertsimas and Tsitsikl [22], which has the advantages of simply description, flexible use, extensive apply and high efficiency. Guimera *et al*. [23] presented a clustering algorithm GA in complex networks based on simulated annealing algorithm, and applies it to the metabolic network. The result shows that this algorithm can quickly find the optimal solution to community detection based on modularity, but it has the deficiencies that the global search capability is poor and the performance is easy to be restrained by parameters.

In community detection algorithm, simulated annealing algorithm can quickly find the optimal solution to community detection which is based on the modularity. But there are some disadvantages of poor global search and the performance can be easily affected by parameter constraints.

Synthesizing the advantages and disadvantages of the above algorithms, based on infomap and Simulated Annealing (SA) algorithm, and using the thought of

Manuscript received April 1, 2015; revised July 21, 2015.
Correspondence author email: yhliu@mail.ccnu.edu.cn
doi:10.12720/jcm.10.7.503-511

optimization to the modularity function, the authors are proposing a novel algorithm Infomap-SA of detecting community.

## II. Relevant Algorithmsand Evaluation Functions of Community Detection

### A. Infomap Algorithm

Infomap algorithm was presented by Rosvall and Bergstroin2008 [24]. This algorithm considers the code length of a random walk of the map as the objective function to be optimized, and detects the network of associations by compressing the information coding. If each node is indicated by the specified binary Huffman coding, any path in the network can be represented with a particular codeword. In order to divide the network, firstly, each community will be coded according to the level of community, and then within each community, nodes will be coded based on the level of node. By integrating these two aspects, the coding of one node is confirmed by community-coding and node-coding. Therefore, the problem of community detection can be replaced by the problem of coding compression, which makes the length of coding shortest. When optimizing the objective function, the algorithm will divide the nodes which connect closely with each other into the same community, because the objective function is the total length of the coding of random walk paths in the network. In this case, the best method to detect community will gain the maximum amount of coding compression. At present, Infomap is considered as one of the most optimal community detection algorithms, but the disadvantage is that the computable complexity of algorithm is complicated.

### B. Simulated Annealing Algorithm

Simulated Annealing (SA) was a probabilistic method proposed by Bertsimas and Tsitsikl for finding the global minimum of a cost function that may possess several local minima [22]. It works by emulating the physical process where a solid is slowly cooled and eventually its structure is “frozen”, which happens at a minimum energy configuration. SA is a generally applicable and easy to implement probabilistic approximation algorithm that is able to produce good solutions for an optimization problem, even if the structure of the problem is unknown. However, there need more theoretical and experimental researches to further assess the potential of the method.

### C. Evaluation Functions of Community Detection

#### 1) Modularity

Some of the evaluation criterions are used to assess the results of community detection obtained by some algorithms. Therefore, in 2004, the modularity $Q$ metric was proposed by Newman and Girvan [9]. It effectively reflects the difference between the actual number of connection and the excepted number of connection in random condition. The equation is as follows:

$$Q = \sum_{s=1}^{K}\left[\frac{m_s}{m} - \left(\frac{d_s}{2m}\right)^2\right] \tag{1}$$

where $Q$ represents the modularity, $K$ is the number of communities in the network, $m$ is the sum of all edges in the whole network, $m_s$ is the sum of all edges in community $s$, $d_s$ is the sum of the degree of all nodes in $s$. The modularity $Q$ metric is used to verify the accuracy of community detection in complex network. It is generally accepted that the value of modularity is higher, the result of community detection is better.

#### 2) Density

The density of one community is the ratio of the total edges in this community and the total edges in network [25], [26]. The equation is as follows:

$$P = \sum_{s=1}^{K}\frac{m_s}{m} \tag{2}$$

where $P$ represents the density, $K$ is the number of communities in the network, $m$ is the sum of all edges in the whole network, $m_s$ is the sum of all edges in community $s$. It can effectively reflect the closeness in internal community. The density $P$ function is used to verify the accuracy of community detection in complex network. It is generally accepted that the value of density is higher, the result of community detection is better.

## III. Realization of the Infomap-SA Algorithm

### A. Algorithm Thought

Firstly, by using the characteristic of random walk coding in Infomap algorithm, the Infomap-SA algorithm names each community with different code, and names each node in every community with recursive process, then compresses the code to gain the community whose length of coding is minimum and the relationship between nodes is the closest. Secondly, combining with the thoughts of modularity optimization, the Infomap-SA algorithm uses the thought of simulated annealing to randomly move the node in the previous community into another one. After this process, the value of modularity is tested, if the result is greater than the previous value, then this new method of community detection is adopted. Otherwise, this new method will be accepted with some probability, until the communities are detected accurately in complex network.

Compared with other community detection algorithms, this proposed algorithm not only can reflect the mutual interdependence and relationship between local communities in global network, but also can optimize the modularity using the thought of simulated annealing with the characteristics of flexibility and high efficiency.

### B. Steps of the Algorithm

The steps of Infomap-SA are given as follows:

**Step 1.** Constructed the set of complex network $G(V, E)$, where $G$ represents the whole network, $V$ is the set of nodes and $E$ is the set of edges. The number of nodes is $n$ ($\alpha = 1, 2, \cdots, n$) and the number of modules is $A$ ($i = 1, 2, \cdots, A$).

**Step 2.** According to Shannon source coding theorem, the entropy $H(\mathcal{P}^i)$ of module coding $i$ is defined as

$$H(\mathcal{P}^i) = -\frac{q_{i\curvearrowright}}{q_{i\curvearrowright} + \sum_{\beta\in i} p_\beta}\log(\frac{q_{i\curvearrowright}}{q_{i\curvearrowright} + \sum_{\beta\in i} p_\beta}) - \sum_{\alpha\in i}\frac{p_\alpha}{q_{i\curvearrowright} + \sum_{\beta\in i} p_\beta}\log(\frac{p_\alpha}{q_{i\curvearrowright} + \sum_{\beta\in i} p_\beta}) \tag{3}$$

where $q_{i\curvearrowright}$ is the probability to exit module $i$, $p_\beta$ is the probability to visit node $\beta$ in module $i$.

**Step 3.** The equation of index coding is defined as

$$H(Q) = -\sum_{i=1}^{A}\frac{q_{i\curvearrowright}}{\sum_{j=1}^{A} q_{j\curvearrowright}}\log(\frac{q_{i\curvearrowright}}{\sum_{j=1}^{A} q_{j\curvearrowright}}) \tag{4}$$

**Step 4.** Put the results generated in step 2 and step 3 into the following equation

$$\begin{aligned} L(G) &= q_{\curvearrowright} H(Q) + \sum_{i=1}^{A} p_{\circlearrowright}^{i} H(\mathcal{P}^i) \\ &= (\sum_{i=1}^{A} q_{i\curvearrowright})\log(\sum_{i=1}^{A} q_{i\curvearrowright}) - 2\sum_{i=1}^{A} q_{i\curvearrowright}\log(q_{i\curvearrowright}) \\ &- \sum_{\alpha=1}^{n} p_\alpha \log(p_\alpha) + \sum_{i=1}^{A}(q_{i\curvearrowright} + \sum_{\alpha\in i} p_\alpha)\log(q_{i\curvearrowright} + \sum_{\alpha\in i} p_\alpha) \end{aligned} \tag{5}$$

where $q_{\curvearrowright} = \sum_{i=1}^{A} q_{i\curvearrowright}$, $p_{\circlearrowright}^{i} = \sum_{\alpha\in i} p_\alpha + q_{i\curvearrowright}$. The minimum coding length $L(G)$ is calculated. If $G$ is an undirected graph, then

$$\begin{aligned} L(G) &= w_{\curvearrowright}\log(w_{\curvearrowright}) - 2\sum_{i=1}^{A} w_{i\curvearrowright}\log(w_{i\curvearrowright}) \\ &- \sum_{\alpha=1}^{n} w_\alpha \log(w_\alpha) + \sum_{i=1}^{A}(w_{i\curvearrowright} + w_i)\log(w_{i\curvearrowright} + w_i) \end{aligned} \tag{6}$$

where $w_i$ is relative weight of the $i$th module, $w_\alpha$ is relative weight of the $\alpha$ th node, $w_i = \sum_{\alpha\in i} w_\alpha$. $w_{i\curvearrowright}$ is the relative weight of connected edges separated from the $i$th module, $w_{\curvearrowright} = \sum_{i=1}^{m} w_{i\curvearrowright}$ is the relative weight between all modules. Relative weight of one node refers to that the weight of the edges connected to it divides double of total weight in the network. In the end, the node with minimum average coding length is divided into the module which the $i$th node belongs to. If the value of equation $L(G)$ is not expected, the steps 2, 3 and4are repeated.

**Step 5.** $Q_{\max}$ is the value of modularity of community detection. The initial temperature is set as $T$, move the node randomly as $dE = Q_{now} - Q_{\max}$, if $dE > 0$, save $Q_{\max}$ as $Q_{now}$, if $Q_{now} < Q_{\max}$, estimate the result of the following equation

$$\exp(dE/T) - random(0...1) \tag{7}$$

if the result of equation (7) is greater than 0, save $Q_{\max}$ as $Q_{now}$. Do iterate at current temperature until the optimal solution is found.

**Step 6.** After calculating the equation $T - \Delta T$, repeat step 5 until the terminal condition is satisfied. Output the optimal result of community detection.

The Infomap-SA algorithm is given as Fig. 1.

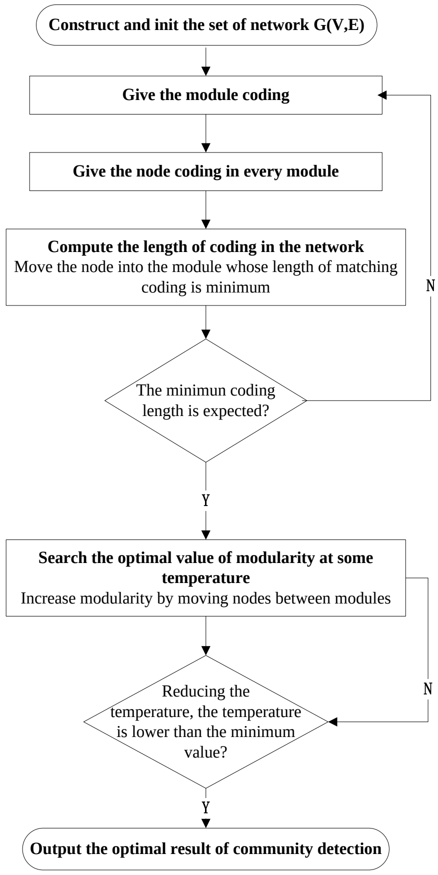



Fig. 1. The flowchart of Infomap-SA algorithm.

## C. Computable Complexity

The calculation process of Infomap-SA algorithm is divided into two steps. Firstly, the communities and nodes are coded respectively by Infomap algorithm, the coding length is minimized, where $n$ is the number of nodes in network, the computable complexity is $O(n^3)$. Secondly, the modularity of the community detection is optimized by simulated annealing algorithm. This method can reduce the previous time complexity into $O(n^2)$. So, the final computable complexity of Infomap-SA is $O(n^2)$.

## IV. Simulations and Analysis

In this paper, some experiments have been done to verify the performance of Infomap-SA algorithm. Linux and C++ is the simulation software environment for this new algorithm. In order to analyze further the performance of Infomap-SA, we compare this algorithm with the Infomap algorithm [27]. The evaluation functions include modularity and density, and the verify networks include real-world networks and generated-networks.

### *A. Tests on Real-World Networks*

To test Infomap-SA, we have performed our experiments over three real-world networks. They are the wellknown Zachary's karate club network [28], bottlenose dolphin network [29] and Les Miserables network [30].

#### *1) Test on Zachary's karate club network*

Karate club network was constructed by Wayne Zachary after observing social interactions between the members of a karate club at an American university [28]. This network includes 34 nodes and 78 edges, in which one node represents a member, and one edge denotes the friendship between any two members.

The result of community detection calculated by Infomap-SA in Karate club network is shown in Fig. 2, the value of modularity is 0.4025, the value of density is 0.8023, and the value of accuracy is 0.9706. In this network, the administrator and the instructor are represented by nodes $v_1$ and $v_{34}$ respectively. Based on the middle black line, this network is divided into two large communities, the right one and left one represent the instructor's faction and the administrator's faction respectively. Furthermore, the left community can also be split into three sub-communities, which are identified by blue, red and green circles. At the same time, the result acquired by this algorithm is identical to the classical paper [11].

#### *2) Test on bottlenose dolphin network*

Bottlenose dolphin network was compiled by Lusseau after seven years of field studies of the dolphins living off Doubtful Sound, New Zealand [29]. This network includes 62 nodes and 159 edges, in which one node represents a dolphin, and one edge denotes the frequent association between any two dolphins.

The result of community detection calculated by Infomap-SA in bottlenose dolphin network is shown in Fig. 3, the value of modularity is 0.5129, the value of density is 0.7624, and the value of accuracy is 0.9839. Based on the middle black line, this network is split into two large communities. Further, the left one also can be divided into five sub-communities denoted by the different colors of nodes. At the same time, the result acquired by this algorithm is identical to the classical paper [11].

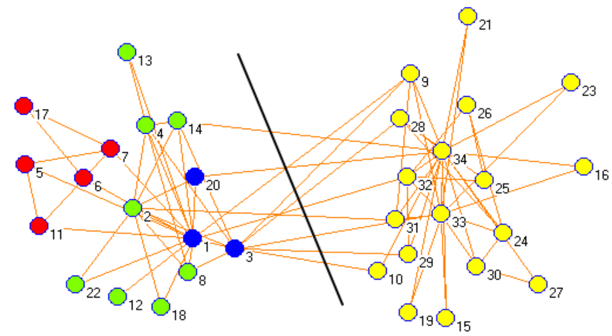


Fig. 2. Communities in Zachary's karate club network.

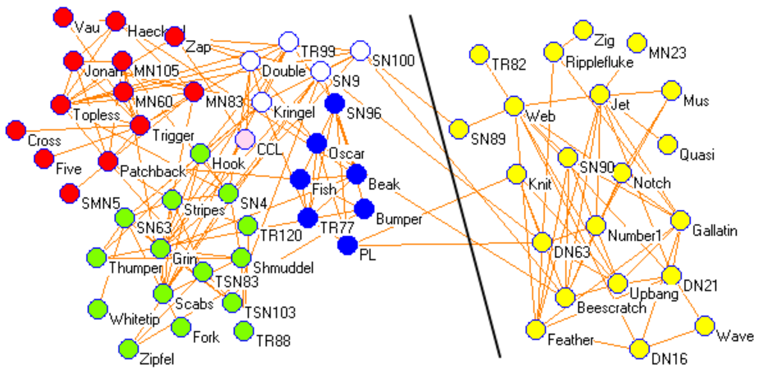



Fig. 3. Communities in Bottlenose dolphin network.

#### *3) Test on les miserables network*

Les Miserables network was compiled by Knuth using the interactions between major characters in Victor Hugo's sprawling novel, Les Miserables [30]. This network includes 77 nodes and 508 edges, in which one node represents a character, and one edge denotes co-appearance of the corresponding characters in one or more scenes.

The result of community detection calculated by Infomap-SA in Les Miserable network is shown in Fig. 4, the number of communities is 11, the value of modularity is 0.4602, the value of density is 0.7402, and the value of accuracy is 0.9481. The communities clearly reflect the subplot structure of the network, unsurprisingly, the protagonist Jean Valjean and his nemesis, the police officer Javert, which are marked by black squares, are central to the network and form the hubs of communities composed of their respective adherents. Other subplots centered on Marius, Cosette, Fantine, and the bishop Myriel, which are marked by black circles, are also picked out by this algorithm. At the same time, the result acquired by this algorithm is identical to the classical paper [11].

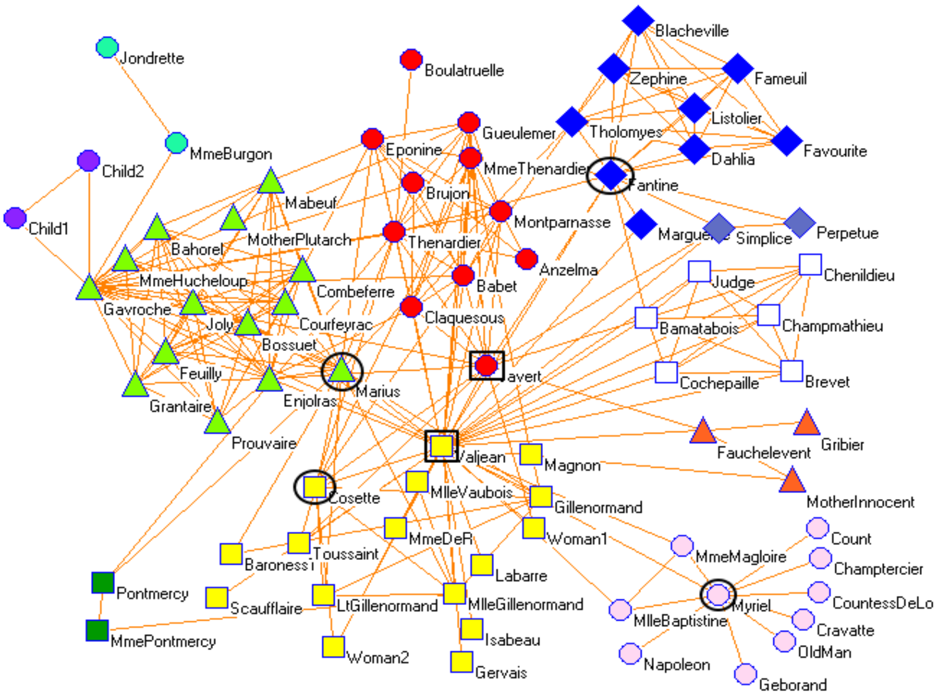



Fig. 4. Communities in les miserable network.

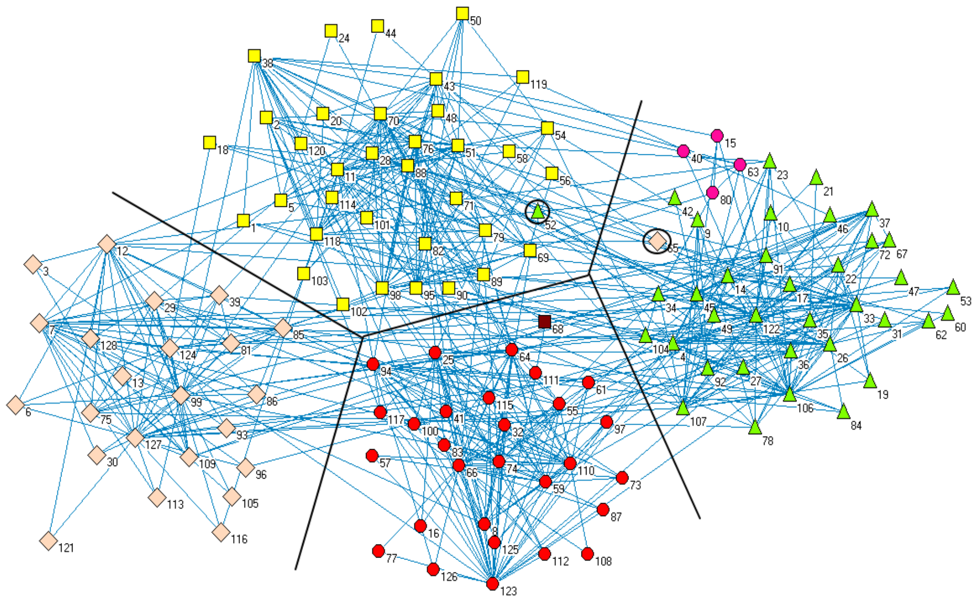


Fig. 5. The network generated by $LFR(128,10,30,1,1,0.2)$.

## B. *Tests on Computer-Generated Networks*

In social networks, the number of nodes and degree of nodes usually satisfy the power-law distribution. Based on LFR-benchmark, generated-networks can be acquired through setting different rules of power-law distribution, and these networks are closer to the real-world networks [31]. Furthermore, some networks with same number of nodes and different parameters can be generated by LFR-benchmark. At the same time, the generated nodes belonging to which community can be already known in advance, so the community structure is certain. This kind of generated-networks is defined as $LFR(N,k,m,\gamma,\phi,\varphi)$, where $N$ is the number of nodes in networks, $k$ is the average degree of nodes, $m$ is the maximum degree of nodes, $\gamma$ is the exponent for the degree distribution, $\phi$ is the exponent for the community size distribution, and $\varphi$ is the mixing parameter which is used to control the ratio between the degree of intra-communities of a node and its total degree. In the Fig. 5, the network generated by $LFR(128,10,30,1,1,0.2)$ is used as the example.

The result of community detection calculated by Infomap-SA in generated-network is shown in Fig.5, the number of communities is 6, the value of modularity is 0.3935, the value of density is 0.7957, and the value of accuracy is 0.9844. Based on the black lines, this network is split into four large communities, which is basically the same to the community structure known in advance generated by LFR-benchmark [31]. But the nodes $v_{52}$ and $v_{65}$ are divided into the wrong communities by Infomap-SA algorithm, which are marked by black circles.

## V. Analysis

### A. Contrastive Analysis of Infomap-SA and Infomap in Real-World Networks

In this paper, modularity and density are calculated by Infomap-SA algorithm and Infomap algorithm based on eight real-world networks, including Zachary's karate club network [29], bottlenose dolphin network [30], American college football network [6], Jazz bands network [32], Les Miserables network [30], Krebs polbooks network [9], Email network of RovirsIVirgili (URV) [33] and the power grid network of the western United States [34]. The contrastive analysis of Infomap-SA and Infomap in different real-world networks is as follows.

The modularity $Q$ metric is used to verify the accuracy of community detection in complex network. It is generally accepted that the value of modularity is higher, the result of community detection is better [9]. In Table I, the number of nodes, the edges and the results of modularity calculated by Infomap-SA and Infomap in different networks are listed. As shown in Table I and Fig. 6, we can see that the values of $Q$ metric calculated by Infomap-SA and Infomap algorithms are high, which indicates that these two algorithms perform well on these eight real-world networks. Obviously, Infomap-SA algorithm performs better than Infomap, and also has higher modularity than Infomap.

TABLE I: THE MODULARITY COMPARISON OF INFOMAP-SA AND INFOMAP.

| Networks | Nodes | Edges | Modularity | |
|---|---|---|---|---|
| | | | Infomap-SA | Infomap |
| Karate | 34 | 78 | 0.4025 | 0.2485 |
| Dolphins | 62 | 159 | 0.5129 | 0.4208 |
| Football | 115 | 613 | 0.5848 | 0.4836 |
| Jazz | 198 | 2742 | 0.4949 | 0.1946 |
| Les Miserables | 77 | 254 | 0.4602 | 0.4082 |
| Pol-books | 105 | 441 | 0.5279 | 0.4546 |
| Email | 1133 | 10903 | 0.4899 | 0.4769 |
| Power | 4941 | 13188 | 0.8269 | 0.8133 |

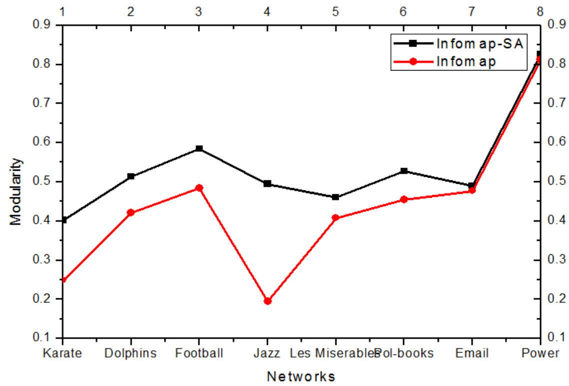



Fig. 6. The modularity comparison chart of Infomap-SA and Infomap.

TABLE II: THE DENSITY COMPARISON OF INFOMAP-SA AND INFOMAP.

| Networks | Nodes | Edges | Density | |
|---|---|---|---|---|
| | | | Infomap-SA | Infomap |
| Karate | 34 | 78 | 0.8023 | 0.8205 |
| Dolphins | 62 | 159 | 0.7624 | 0.7736 |
| Football | 115 | 613 | 0.3268 | 0.1373 |
| Jazz | 198 | 2742 | 0.4055 | 0.1809 |
| Les Miserables | 77 | 254 | 0.7402 | 0.7244 |
| Pol-books | 105 | 441 | 0.8571 | 0.8549 |
| Email | 1133 | 10903 | 0.5874 | 0.5735 |
| Power | 4941 | 13188 | 0.8287 | 0.8174 |

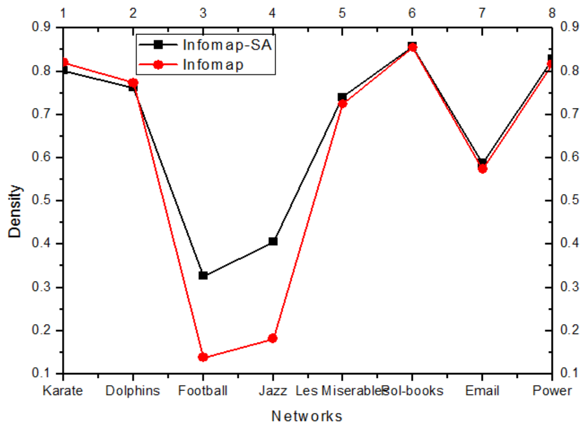



Fig. 7. The density comparison chart of Infomap-SA and Infomap.

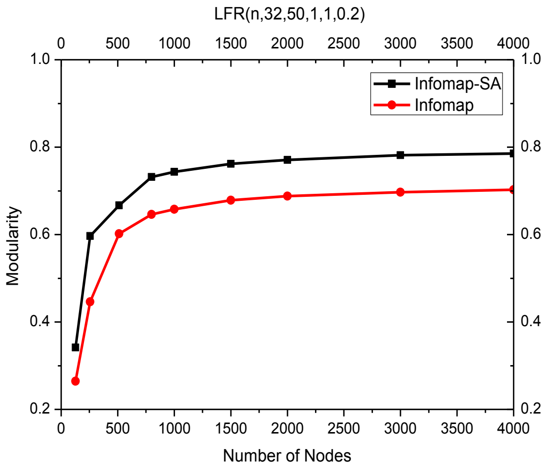
The density $P$ function is also used to verify the effectiveness of community detection in complex network. It is generally accepted that the value of density is higher, and the result of community detection is better [25], [26]. In Table II, the number of nodes, the edges and the results of density calculated by Infomap-SA and Infomap in different networks are listed. As shown in Table II and Fig. 7, we can see that the values of $P$ function obtained by Infomap-SA and Infomap algorithms are high, which shows that Infomap-SA and Infomap perform well on these eight real-world networks. Obviously, the values calculated by Infomap-SA are just lower than Infomap in karate network and dolphin network, but values are higher than Infomap in other four networks, which demonstrates that Infomap-SA algorithm performs better than Infomap with the increasing of the network scale.

TABLE III: THE MODULARITY COMPARISON OF INFOMAP-SA AND INFOMAP.

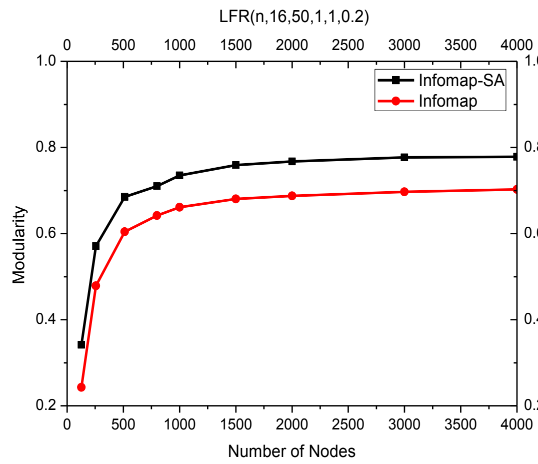
| Number of Nodes | *LFR*(*n*,32,50,1,1,0.2) | | *LFR*(*n*,16,50,1,1,0.2) | | *LFR*(*n*,10,30,1,1,0.2) | |
|---|---|---|---|---|---|---|
| | Infomap-SA | Infomap | Infomap-SA | Infomap | Infomap-SA | Infomap |
| 4000 | 0.7855 | 0.7027 | 0.7783 | 0.7024 | 0.7853 | 0.7030 |
| 3000 | 0.7818 | 0.6969 | 0.777 | 0.6970 | 0.7823 | 0.6989 |
| 2000 | 0.7707 | 0.6878 | 0.7675 | 0.6873 | 0.772 | 0.6902 |
| 1500 | 0.762 | 0.6788 | 0.7589 | 0.6807 | 0.7623 | 0.6820 |
| 1000 | 0.7436 | 0.6577 | 0.735 | 0.6613 | 0.7401 | 0.6566 |
| 800 | 0.732 | 0.6460 | 0.7102 | 0.6419 | 07342 | 0.6531 |
| 512 | 0.667 | 0.6018 | 0.6854 | 0.6042 | 0.6994 | 0.6276 |
| 256 | 0.5969 | 0.4463 | 0.5709 | 0.4788 | 0.6347 | 0.4962 |
| 128 | 0.3415 | 0.2643 | 0.3417 | 0.2429 | 0.3935 | 0.3029 |

TABLE IV: THE DENSITY COMPARISON OF INFOMAP-SA AND INFOMAP.

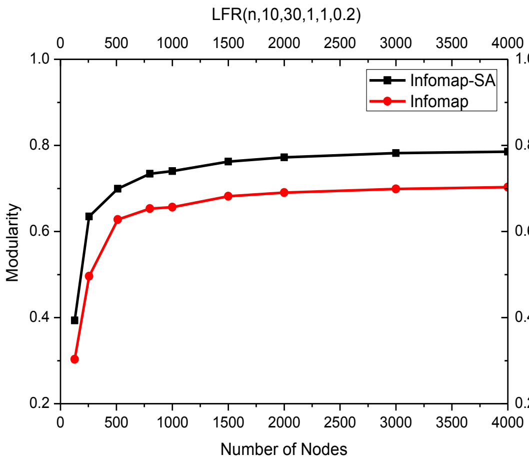
| Number of Nodes | *LFR*(*n*,32,50,1,1,0.2) | | *LFR*(*n*,16,50,1,1,0.2) | | *LFR*(*n*,10,30,1,1,0.2) | |
|---|---|---|---|---|---|---|
| | Infomap-SA | Infomap | Infomap-SA | Infomap | Infomap-SA | Infomap |
| 4000 | 0.8 | 0.7811 | 0.7935 | 0.7710 | 0.7984 | 0.7705 |
| 3000 | 0.8 | 0.7806 | 0.7972 | 0.7707 | 0.7995 | 0.7701 |
| 2000 | 0.8 | 0.7807 | 0.7985 | 0.7700 | 0.8 | 0.7704 |
| 1500 | 0.7994 | 0.7804 | 0.8 | 0.7621 | 0.8 | 0.7605 |
| 1000 | 0.8 | 0.7806 | 0.7963 | 0.7607 | 0.7978 | 0.7610 |
| 800 | 0.8 | 0.7811 | 0.8 | 0.7601 | 0.7972 | 0.7606 |
| 512 | 0.8 | 0.7797 | 0.8013 | 0.7618 | 0.7981 | 0.7606 |
| 256 | 0.7994 | 0.7712 | 0.8085 | 0.7642 | 0.7955 | 0.7632 |
| 128 | 0.7966 | 0.7728 | 0.7983 | 0.7622 | 0.7957 | 0.7642 |




(a)




(b)

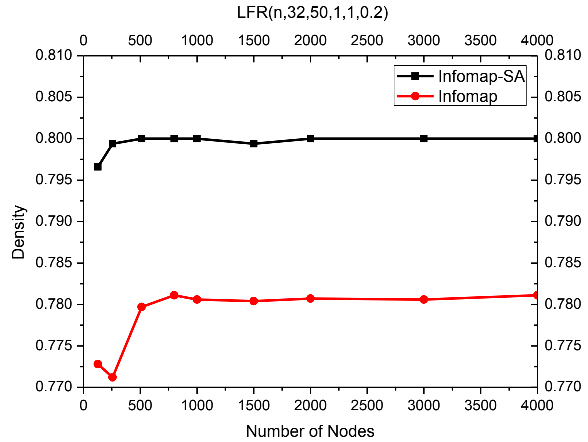



(c)

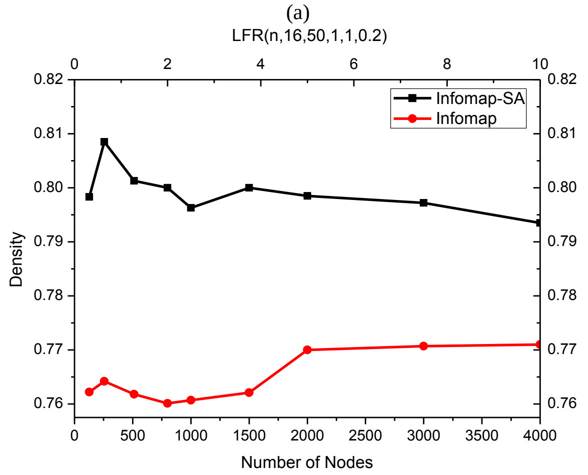
Fig. 8. (Color online)The modularity comparison chart of Infomap-SA and Infomap: (a) networks generated by $LFR(n,32,50,1,1,0.2)$ . (b) networks generated by $LFR(n,16,50,1,1,0.2)$ . (c) networks generated by $LFR(n,10,30,1,1,0.2)$ .

### B. Contrastive Analysis of Infomap-SA and Infomapin Different Computer-Generated Networks

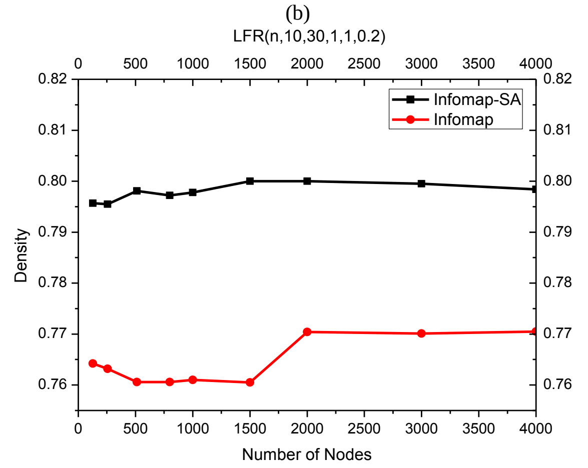
As a further test on the Infomap - SA algorithm, we apply it to the computer-generated networks based on LFR-benchmark [35], which have different topological properties than the real-world ones. Here we use 27 generated networks to perform Infomap-SA algorithm. These networks are generated by $LFR(N,k,m,\gamma,\phi,\varphi)$ with different parameters, in which the number of nodes $N$ is from 128 to 4000, the average degree $k$ is 10, 16 or 32, the maximum degree of nodes $m$ is 30 or 50, the exponent for the degree distribution $\gamma$ is 1, the exponent for the community size distribution $\phi$ is 1, the mixing parameter $\varphi$ is 0.2.




(a)




(b)




(c)

Fig. 9. (Color online) The density comparison chart of Infomap-SA and Infomap: (a) networks generated by $LFR(n,32,50,1,1,0.2)$, (b) networks generated by $LFR(n,16,50,1,1,0.2)$, and (c) networks generated by $LFR(n,10,30,1,1,0.2)$.

Based on Table III and Fig. 8, we can see that the values of modularity calculated by Infomap-SA and Infomap algorithms are increasing gradually with the increase of network scale. Obviously, the former performs better than the latter, and also have higher modularity than the latter.

According to Table IV and Fig. 9, we can see that the values of density acquired by Infomap-SA and Infomap algorithms are high. Obviously, the former performs better than the latter, and also have higher density than the latter. The result of detecting communities calculated by Infomap-SA is better than Infomap.

So, based on the comprehensive analysis above, it is obvious that the Infomap-SA algorithm performs well in real-world networks and generated networks. And the values of modularity and density calculated by this algorithm are higher than Infomap algorithm.

## VI. Conclusions

The identification of community structure in complex networks is of great interest because they often reveal unknown relations among nodes and provide useful information for unknown things. In this paper, based on Infomap algorithm, and integrating the thought of simulated annealing to optimize the modularity function, the authors are proposing a new algorithm Infomap-SA of community detection. The performance of this algorithm is tested on six real-world networks, as well as on a set of computer-generated networks by LFR-benchmark. Through further contrastive analyses, the experimental results reveal that the Infomap-SA algorithm can identify the communities accurately and efficiently, and have higher values of modularity and density and lower computable complexity than Infomap algorithm. However, there also exist some problems in this algorithm. For example, it cannot overcome the problem of poor global searching capability in simulated annealing algorithm. Therefore, our future researches will aim at identifying communities accurately by using the methods with more stable global solutions.

## References

[1] D. J. Watts, "A twenty-first century science," *Nature*, vol. 445, pp. 489, Feb., 2007.

[2] J. C. Cai and B. M. Yu, "A discussion of the effect of tortuosity on the capillary imbibition in porous media," *Transport in Porous Media*, vol. 89, no. 2, pp. 251-263, July 2011.

[3] A. E. Motter, S. A. Myers, M. Anghel, and T. Nishikawa, "Spontaneous synchrony inpower-grid networks," *Nature Physics*, vol. 9, no. 3, pp. 191–197, Feb. 2013.

[4] D. Lazer, A. Pentland, L. Adamic, *et al.*, "Life in the network: The coming age of computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, Feb. 2009.

[5] F. Y. Wang, D. Zeng, K. M. Carley, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 79–83, Mar. 2007.

[6] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *The National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, Dec. 2002.

[7] J. C. Cai, E. Perfect, C. L. Cheng, and X. Y. Hu, "Generalized modeling of spontaneous imbibition based on hagen-poiseuille flow in tortuous capillaries with variably shaped apertures," *Langmuir*, vol. 30, no. 18, pp. 5142-5151, May 2014.

[8] E. Gegov, M. N. Postorino, M. Atherton, and F. Gobet, "Community structure detection in the evolution of the united

states airport network," *Advances in Complex Systems*, vol. 16, no. 1, pp. 1350003, Mar. 2013.

[9] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, Feb. 2006.

[10] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3, pp. 75–174, December 2009.

[11] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, pp. 026113, Aug. 2004.

[12] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, pp. 066133, Sep. 2004.

[13] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, pp. 066111, Aug. 2004.

[14] B. H. Good, Y. A. D. Montjoye, and A. Clauset, "Performance of modularity maximization in practical contexts," *Physical Review E*, vol. 81, pp. 046106, Apr. 2010.

[15] A. Lancichinetti and S. Fortunato, "Limits of modularity maximization in community detection," *Physical Review E*, vol. 84, pp. 066122, Dec. 2011.

[16] J. Z. Jin, Y. H. Liu, L. T. Yang, *et al*., "An Efficient detecting communities algorithm with self-adapted fuzzy C-Means clustering in complex networks," *in Proc. IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, Shanghai, 2012, pp. 1988-1993.

[17] M. G. Gong, J. Liu, L. J. Ma, *et al*., "Novel heuristic density-based method for community detection in networks," *Physica A*, vol. 403, pp. 71-84, Feb. 2014.

[18] M. Crampes and M. Plantie, "A unified community detection, visualization and analysis method,"*Advances in Complex Systems*, vol. 17, no. 1, pp. 1450001, Mar. 2014.

[19] Y. H. Liu, J. Z. Jin, Y. Zhang, *et al.*, "A new clustering algorithm based on data field in complex networks," *The Journal of Supercomputing*, vol. 67, no. 3, pp. 723-737, Mar. 2014.

[20] C. H. Mu, Y. Liu, Y. Liu, J. S. Wu, and L. C. Jiao, "Two-stage algorithm using influence coefficient for detecting the hierarchical, non-overlapping and overlapping community structure," *Physica A*, vol. 408, pp. 47-61, Apr. 2014.

[21] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure,"*Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118-1123, Dec. 2008.

[22] D. Bertsimas and J. Tsitsiklis, "Simulated annealing,"*Statistical Science*, vol. 8, no. 1, pp. 10-15, Mar. 1993.

[23] R. Guimera and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, pp. 895−900, Feb. 2005.

[24] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *The European Physical Journal*, vol. 178, no. 1, pp. 13-23, Nov. 2009.

[25] A. Lancichinetti, M. Kivela, *et al.*, "Characterizing the community structure of complex networks," *Plos One*, vol. 5, no. 8, pp. e11976, Aug. 2010.

[26] A. Clauset, "Finding local community structure in networks," *Physical Review E*, vol. 72, pp. 026132, Feb. 2008.

[27] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis," *Physical Review E*, vol. 80, pp. 056117, Sep. 2009.

[28] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452-473, May 1977.

[29] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, "The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations," *Behavioral Ecology and Sociobiology*, vol. 54, no. 4, pp. 396-405, June 2003.

[30] D. E. Knuth, "The stanford graph base: A platform for combinatorial computing," *Addison-Wesley*, *Reading*, MA, 1993.

[31] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical Review E*, vol. 78, no. 4, pp. 046110, Oct. 2008.

[32] P. M. Gleiser and L. Danon, "Community structure in Jazz," *Advances in Complex Systems*, vol. 6, no. 4, pp. 1-12, July2003.

[33] R. Guimera, L. Danon, A. D. Guilera, F. Giralt, and A. Arenas, "The real communication network behind the formal chart: Community structure in organizations," *Journal of Economic Behavior & Organizatio*n, vol. 61, no. 4, pp. 653-667, July 2006.

[34] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440-442, June 1998.

[35] J. Z. Ji, X. J. Song, C. N. Liu, and X. Z. Zhang, "Ant colony clustering with fitness perception and pheromone diffusion for community detection in complex networks," *Physica A*, vol. 392, pp. 3260-3272, Apr. 2013.

**Fang Hu** was born in Hubei Province, China, in 1981. She received the B.S. degree from the School of Computer, Central China Normal University (CCNU) in 2004, majoring in computer science and technology, and the M.S. degree from the Computer School of Wuhan University in 2008, majoring in computer software and theory. She is currently pursuing the Ph.D. degree with the School of Computer Science, CCNU. Her research interests include network communication and complex networks.

**Yu-Hua Liu** was born in Hubei Province, China, in 1951. She received her Ph.D. degree from the School of Computer Science & Technology, Huazhong University of Science and Technology, majoring in computer architecture. Currently, she is employed as the Professor and Doctoral Supervisor at the School of Computer Science, Central China Normal University, where she is responsible for teaching and research activities. Her research interests include Computer Networks and Communications, Wireless Networks Technology, Complex Networks Theory, etc.