Physica A 526 (2019) 121070

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Complex network community detection method by improved density peaks model

Zheng-Hong Deng [a,*], Hong-Hai Qiao [a], Ming-Yu Gao [b], Qun Song [a], Li Gao [c]

[a] School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
[b] School of electronic information, Hangzhou Dianzi University, Hangzhou 310026, China
[c] School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

HIGHLIGHTS

- Research on complex network community detection algorithm based on improved density peaks model.
- The algorithm obtains key nodes by density peaks model with threshold conditions.
- Community attribute of unstable nodes is corrected iteratively by neighbor estimation principle.
- The accuracy of community detection is improved.

ARTICLE INFO

Article history:
Received 10 September 2018
Received in revised form 3 March 2019
Available online 12 April 2019

Keywords:
Community detection
Density peaks model
Neighbor estimation principle
Modularity
Normalized mutual information

ABSTRACT

Searching for key nodes of complex network and clustering communities is an important and practical solution in community detection algorithms. According to above-mentioned notion, we proposed a novel complex community detection method by improved density peaks model, called IDPM. Primarily, the composite similarity is acquired by normalizing Jaccard and shortest path feature. Secondarily, threshold condition is acceded to density peaks model to obtain key nodes of network. Tertiarily, non-key nodes are assigned into groups and form base communities. The attribute of unstable nodes are corrected iteratively until the entire network is stable by neighbor estimation principle. The simulation experiments prove that the IDPM can obtain better effects of community detection on synthesize and real-world complex network.

1. Introduction

Community detection algorithms are important component of complex network theory. There is no standard for the strict definition of community detection. Generally, complex network is simplified into several communities, and nodes with common or similar features are assigned to the same community. In the real world, a large number of practical cases can be considered as complex network structures, such as mobile communication networks, bioinformatics [1], power system networks [2,3], Internet, social network [4], information systems [5,6], physics [7]. If complex network structure is clear and concise, the performance analysis of network can achieve optimal results or balance status quickly. Therefore, community detection algorithms are very meaningful to theoretical and practical application.

* Corresponding author.
E-mail address: dthree@nwpu.edu.cn (Z.-H. Deng).

https://doi.org/10.1016/j.physa.2019.121070
0378-4371/

1.1. Relate research directions

1.1.1. Hierarchical structure algorithm

Girvan and Newman [8–10] proposed a community structure detection algorithm (CNM). The similarity measure or connection strength between nodes is regarded as the hierarchical structure features. Nodes are allocated or redistributed into different communities by optimal modularity evaluation criterion.

1.1.2. Genetic algorithms

Shi et al. [11–13] proposed a community detection method based on genetic algorithm (GA). A framework of genetic algorithm for community detection is established, including population (community) initialization, modularity evaluation, nodes and populations updating methods, mutation and crossover operator rules. It uses modularity as the fitness function to distinguish different community attributes or labels by adjacent coding mode. Pizzuti et al. [14] proposed a multi-objective genetic algorithm for community detection (MOGA-Net). According to different hierarchical structure levels, MOGA-Net obtains the optimal community detection results by modularity objective function. Gong et al. [15] proposed a community detection method by improved genetics algorithm (Meme-Net). Parameters are selected to complete the division of network communities accurately by using the uphill strategy and optimization function.

1.1.3. Multi-objective evolutionary algorithms

Shi et al. [16] proposed a multi-objective evolutionary community detection algorithm (MOCD). It establishes a framework of multi-objective community detection to obtain the optimal solution by comparing different objective functions. Gong et al. [17] proposed a community detection method by multi-objective evolutionary algorithm with decomposition (MOEAD-Net). It can increase the internal density of nodes and decrease the external density. A series of optimization solutions are provided to classify nodes into several portions. Zou et al. [18] proposed a discrete multi-objective backtracking search optimization community detection algorithm (MODBSA). The updating rule is redesigned by complex network topology. Community detection results are completed reasonable by negative ratio objective functions.

1.1.4. Label propagation algorithms

Raghavan et al. [19] proposed a community detection algorithm based on label propagation (LPA). The core idea of LPA is to verdict community label of each node by its neighbor nodes and corresponding evaluation methods. Then, community labels are revised continuously until the cut-off conditions are satisfied. LPA is a fast unsupervised parameter clustering algorithm, which can adapt to different types or scale of complex network. Lou et al. [20] proposed an improved label propagation algorithm with weighted coherent neighbor propinquity (CNPE-LPA). It proposed an adaptive weighting method by using entropy information principle. During the update process, the labels of nodes are changed by maximum weighted propinquity.

1.2. Density peaks model

Scholars [21–25] proposed some clustering algorithms by density peaks model. In database, the density peaks model is usually used to select key data packets. Subsequent scholars began to use density peaks model to search for key nodes in the network. Cheng et al. [26] proposed a community detection method by density sequence tree. The core nodes are selected by density peaks model and non-core nodes are allocated according to the access sequence tree.

The core idea of clustering algorithm based on density peaks model is to traverse network and select key nodes by comparing similarity between nodes. The non-key nodes are assigned by relevant allocation methods until the final community detection is completed. The density peaks clustering model can select key nodes of network accurately and adapt to complex network with different structure types or scales. The main process of density peak model is as follows:

(1) Density ρ. Density ρ denotes that the accumulation of discrimination between nodes, which is calculated as shown in Eqs. (1) and (2).

$$\chi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{others.} \end{cases} \tag{1}$$

$$\rho_i = \sum_{j}^{N} \chi(d_{ij} - d_c) \tag{2}$$

where d_{ij} is the similarity distance node n_i and node n_j. d_c is the cut-off threshold. N is the network nodes number.

(2) Minimum distance δ. Each node compares with density ρ of other nodes in the network. The minimum or maximum similarity distance is selected as minimum distance δ, as shown in Eq. (3).

$$\delta_i = \begin{cases} \max\limits_{j}^{N}(d_{ij}), & \text{if } \rho_i = \max\limits_{k}^{N}(\rho_k) \\ \min\limits_{j}^{N}(d_{ij}), & \text{if } \rho_i < \rho_j \end{cases} \tag{3}$$

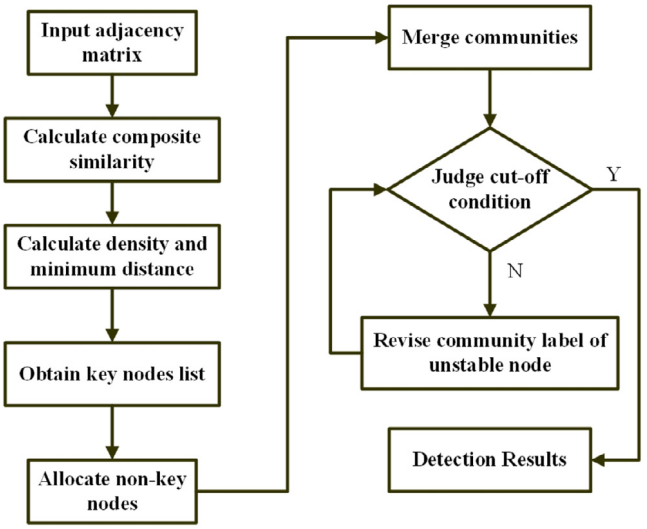


Fig. 1. The framework flowchart of IDPM.

The density ρ and minimum distance δ are sorted as horizontal and vertical coordinates. Nodes with both large abscissa and ordinate coordinates are selected as key nodes.

2. Method

In this paper, a complex network community detection method by improved density peak model is proposed, called IDPM. IDPM method is mainly divided into three main steps and the framework flowchart of IDPM is shown as in Fig. 1.

(1) The calculation of composite similarity. In this paper, the Jaccard index and shortest path information are combined to form composite similarity.

(2) The selection of key nodes. On the basis of step 1, the density and minimum distance of each node are calculated by improved density peaks model. Nodes with large ρ and δ are selected into key nodes list.

(3) Iterative clustering of communities. Non-key nodes are allocated into serial groups and form basic communities. Furthermore, non-basic communities are merged into basic communities by community similarity. Finally, each node should verify own stability according to its neighbor status. Nodes with large diversity are selected to revise theirs community attributes continuously by neighbor estimation principle until the whole network is stable.

2.1. Similarity

Similarity is mainly used to reflect the approximation extent between nodes. Generally, if the larger similarity between nodes is, the closer features, performance and relationships are. The complexity of Jaccard index [26] is very low and can express the extent of approximation between nodes intuitively. Therefore, some algorithms [21–23] use the Jaccard distance to measure the similarity between nodes and establish Jaccard matrix as shown in Eq. (4).

$$J_{ij}^{*} = \frac{M_i \cap M_j}{M_i \cup M_j} \tag{4}$$

where M_i and M_j denote the feature vector of node n_i and node n_j in the adjacency matrix M, respectively. The Jaccard feature J_{ij} is normalized as $J_{ij} = J_{ij}^{*}/J_{max}$.

In the article [26], the relevant metric distances are usually also used to calculate the similarity, such as Salton and LHN et al. However, the expression information of feature distance is too single to reflect the authentic correlation between nodes completely. There are some extreme phenomena that some nodes belonging to different communities still have larger Jaccard feature distance, which will cause adverse effects in subsequent clustering process.

The shortest path information can also convey the connection relation between nodes directly. Therefore, shortest path feature based on adjacency matrix is used to measure the similarity between nodes.

The shortest path matrix P between nodes can be obtained by pairing heap optimal method [27]. The shortest path feature D_{ij} between node n_i and node n_j is transformed and normalized by Eq. (5).

$$D_{ij} = \frac{P_{min}}{P_{ij}} \tag{5}$$

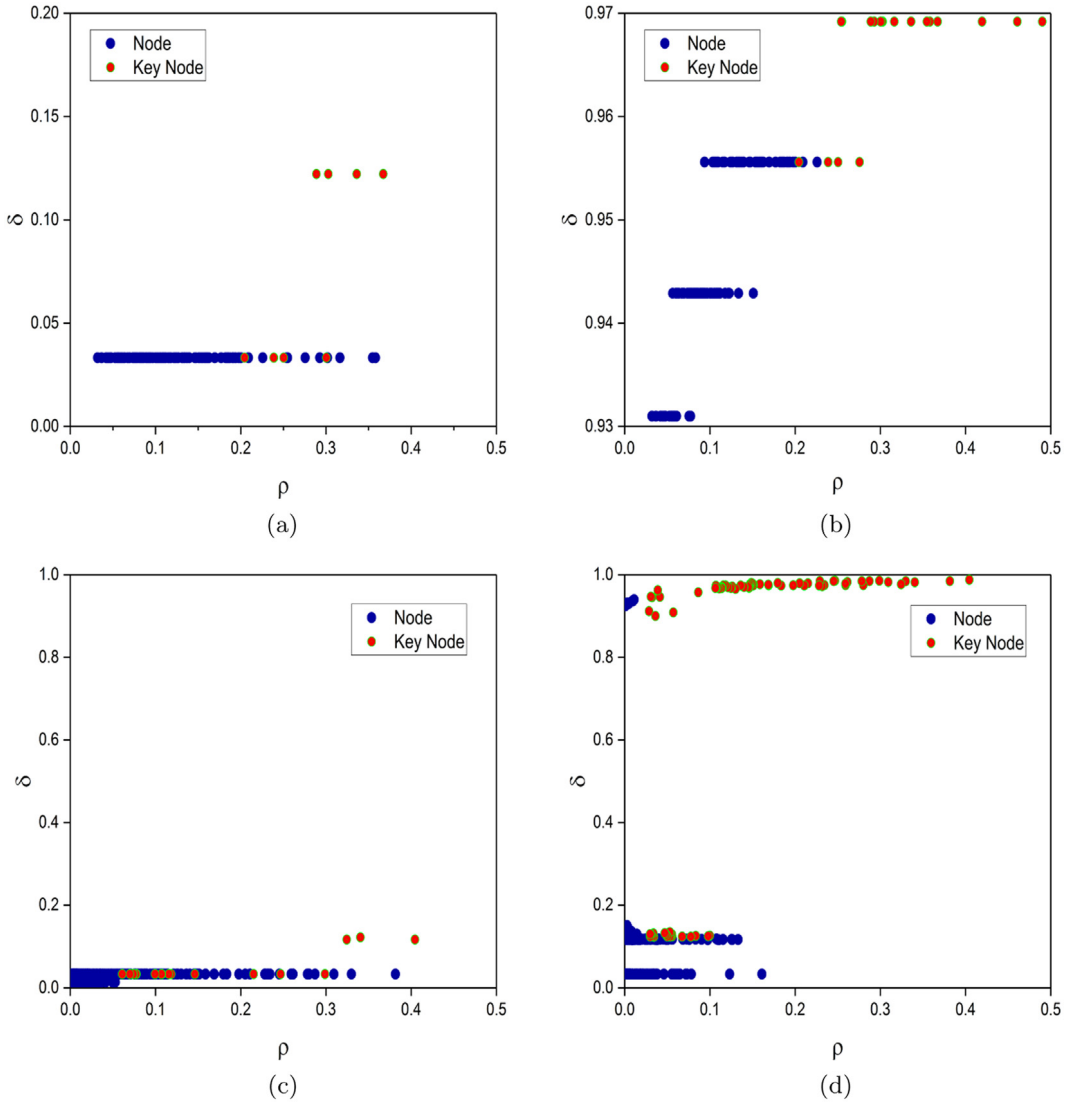


Fig. 2. Schematic diagram of key nodes (red nodes); (a) and (c) are the detection results on GN and LFR network for density peak model without threshold condition. (b) and (d) are the detection results on GN and LFR network by proposed method.

It is possible to ignore the influence of surrounding nodes on approximation extent by shortest path feature. Therefore, according to the merits of Jaccard index and shortest path information, the composite similarity distance is set up as shown in Eq. (6).

$$S_{ij} = (1 - \alpha)D_{ij} + \alpha J_{ij} \tag{6}$$

Composite similarity can be adjusted by parameter α. The composite similarity can reflect the approximate correlation between nodes reasonably, especially for network with complex connectivity structure.

2.2. The selection of key nodes

The issue of community detection is more intractable than common clustering application. Therefore, the extraction model of key nodes should follow three principles: (a) sufficient key nodes; (b) key nodes with representative; (c) large discrimination between key and non-key nodes. The traditional density peaks model does not reflect above criteria well. The selection results of key nodes (red nodes) are shown in Fig. 2(a) and (c).

In the paper, an improved density peak model is proposed to choose the key nodes of network, which is divided into two steps.

(1) Calculation of density ρ. The density will affect the effect of subsequent clustering process directly, and Eq. (7) is chose to compute the density of node n_i.

$$\rho_i^* = \sum_j^N exp(\frac{S_{ij}}{S_c})^2 \tag{7}$$

where S_c is the adjustment parameter. The density ρ_i is normalized as $\rho_i = \rho_i^*/\rho_{max}$. Parameter S_c can be adjusted to change the discrimination between nodes in order to select key nodes accurately.

(2) Density peaks model with threshold condition. The traditional model traverses all nodes of network by comparing density ρ. But most of nodes have small minimum distance δ, which have adverse effects on distinguishing between key and non-key nodes. So we propose an improved strategy that the density threshold condition is introduced into density peaks model.

When the density calculation is completed, nodes' ρ are sorted in descending order. Choosing N_{ckn} nodes with large ρ into candidate list L_ρ. The candidate key nodes number N_{ckn} usually is the 10 $\sim$ 30% of total nodes number on the network. The density threshold ε is determined by the mean deviation of candidate nodes, as shown in Eq. (8).

$$\varepsilon = \frac{\beta}{N_{ckn}} \sum_m^{N_{ckn}} |\rho_m - \overline{\rho}| \tag{8}$$

where $\overline{\rho}$ is the density average of candidate key nodes.

During traversal process, any node n_j is compared with node n_i by threshold condition $\rho_i < \rho_j - \varepsilon$. Furthermore, minimum distance δ_i is obtained as shown in Eq. (9).

$$\delta_i = \begin{cases} \max\limits_j^N(S_{ij}), & \text{if } \rho_i = \max\limits_k^N(\rho_k - \varepsilon) \\ \min\limits_j^N(S_{ij}), & \text{if } \rho_i < \rho_j - \varepsilon \end{cases} \tag{9}$$

Nodes with large $\gamma = \rho\delta$ are usually regarded as key nodes by traditional method, which leads to the phenomena of excessive selection. Some nodes with small ρ or δ are misclassified into key nodes list L_K. Therefore, we sort in descending order. Choosing N_{ckn} nodes with large δ into candidate list L_δ. Those nodes with large ρ and δ are selected as key nodes in the interval of list and simultaneously.

The improved density peaks model allows large of nodes to choose maximum similarity as their minimum distance δ by using threshold conditions. It has positive impact on the selection of key nodes. (a) list L_δ and L_ρ have sufficient same nodes, which increase key nodes number. (b) The average δ of key nodes is raised to enlarge discrimination between key and non-key nodes. (c) Due to threshold ε being decided by candidate nodes with large ρ, nodes with representative are absorbed into list L_K and reduce the error rate of selection.

Therefore, key nodes list L_K can be obtained by improved density peaks model reasonably. The selection results of key nodes (red nodes) are shown in Fig. 2(b) and (d). It is propitious to the allocation of non-key nodes and communities merge in subsequent process. The pseudo code of key nodes selection is shown in Table 1.

2.3. Iterative clustering of communities

The establishment of key nodes list determines community distribution roughly. Each key node is selected as the core of a group. Subsequent community detection is completed by iterative clustering of communities. In this paper, the idea of iterative clustering is as follow: non-key nodes are divided into several groups. Then, groups will be integrated into various communities by the community similarity. Finally, community attributes of unstable nodes are continuously modified until the network is stable. The clustering content is mainly divided into four subordinate contents.

(1) The allocation of non-key nodes. Non-key nodes are assigned into group to which key node with maximum similarity belongs. Node n_v is selected as the core of group G_v. If non-key node n_m and key node n_v have the maximum similarity S_{mv}, and node n_m is assigned to group G_v as shown in Eq. (10).

$$n_m \in G_v \qquad if \quad S_{mv} = \max_z^{N_{ckn}}(S_{mz}) \tag{10}$$

Non-key nodes are allocated and form N_{ckn} small-scale groups. The maximum community number is determined as $C_{max} = \sqrt{E}/2$ by network edge E. The optimal community number will be detected in the interval by modularity.

(2) The merge of communities. Small-scale groups are sorted in descending order by internal nodes number. The groups with large group scale are selected as basic communities. The other groups are regarded as non-basic communities. Then, non-basic communities are merged into basic communities by community similarity.

The community similarity dis_{ij} is calculated between basic community C_i and non-basic community C_j as shown in Eqs. (11)–(14).

$$dis_1 = \frac{2}{N_{C_i}(N_{C_i} - 1)} \sum_{k_1, l_1 \in C_i} S_{k_1 l_1} \tag{11}$$

Table 1
The pseudo code of key nodes selection.

Input: M, N_{ckn}, S_c
Step1: for all i do for all j do calculateD_{ij} ,J_{ij} and S_{ij} end end
Step2: for all i do calculateρ_i end
Step3: sortρ in descending order, and select candidate listL_ρ
Step4: calculate density thresholdε
Step5: for all i do for all j do calculateδ_i by threshold condition end end
Step6: sortδ in descending order, and select candidate listL_δ
Output: Key nodes List L_K

$$dis_2 = \frac{2}{N_{C_j}(N_{C_j}-1)} \sum_{k_2.l_2 \in C_j} S_{k_2 l_2} \tag{12}$$

$$dis_3 = \frac{1}{N_{C_i} N_{C_j}} \sum_{k_3 \in C_i, l_3 \in C_j} S_{k_3 l_3} \tag{13}$$

$$dis_{ij} = |(dis_1 + dis_2)/2 - dis_3| \tag{14}$$

where N_{C_i} and N_{C_j} are the internal nodes number of community C_i and community C_j.

Assuming basic communities $C = \{C_1, C_2 \ldots C_K\}$. Each non-basic community selects the minimum similarity of basic community as its merged community. If dis_{ij} is the minimum similarity, community C_j is merged into basic community C_i as shown in Eq. (15).

$$C_j \in C_i \quad if \quad dis_{ij} = \min_h^K (dis_{hj}) \tag{15}$$

(3) Redistribution. After merge of communities, most of nodes can be assigned to communities correctly. However, there still are a few misclassified nodes, which have adverse influence on clustering accuracy. Therefore, nodes with large diversity are selected in each community and modify their community attributes.

Neighbor estimation principle is used to redistribute unstable nodes. The main content of neighbor estimation principle contains the neighbor feature vectors and community attributes modification. The neighbor feature vector $f_{n_i} = \{f_{n_i}^{C_1}, f_{n_i}^{C_2} \ldots f_{n_i}^{C_K}\}$ between node n_i and communities is shown in Eq. (16). $f_{n_i}^{C_j}$ is the proportion of node n_i's neighbor nodes in community C_j, which reflects the probability of node n_i belonging to community C_j.

$$f_{n_i}^{C_j} = \frac{Nerb_{n_i}^{C_j}}{N_{C_j}} \tag{16}$$

where $Nerb_{n_i}^{C_j}$ is the adjacent nodes number in community C_j.

Each node's status is verified by its neighbor feature vector. If vector f_{n_i}'s maximum community attribute is the current attribute of node n_i, and is regarded as stable node of network. Otherwise, belongs to unstable(diversity) node and its community attribute should be revised by Eq. (17). If community C_n is the attribute of vector f_{n_i} maximum element, and node n_i is assigned to community C_n.

$$n_i \in C_n \quad if \quad f_{n_i}^{C_n} = \max_h^K (f_{n_i}^{C_h}) \tag{17}$$

Due to the complexity of network, it is necessary to compute the feature vector f iteratively and modify attributes of nodes with large diversity to achieve network stable status.

(4) Iterative condition. The entropy information index of network [27,28] can be used to estimate the stability as shown in Eqs. (18)–(19). In iterative process, entropy information by shortest path is not changed any more, which denotes that

Table 2
The pseudo code of iterative communities clustering.

Input: Key nodes List L_K
Step1: select key nodes as the core of groups Step2: for all i do allocate non-key nodes n_i into groups end Step3: sort groups in descending order by group size Step4: select$K$basic communities and $N_{ckn}-K$non-basic communities Step5: for $i=1:K$ for $j=N_{ckn}-K:N_{ckn}$ calculate community similarity by Eq(11)-(14) if condition of Eq(15)is satisfied $C_j$is merged into C_i end end end Step6: while flag for all i do establish feature vector C_i end choose unstable nodes and reallocate them by Eq(17) calculate entropyI_t by Eq(18)-(19) if $\|I_t-I_{t-1}\|=0$, iterative calculation end and flag=0, return Step7 otherwise, $t=t+1$ and return to Step 6 end Step7: select the optimal result
Output: Final results

the community attribute between node and its neighbor nodes are consistent or stable. If the condition $|I_t - I_{t-1}| = 0$ is satisfied, iterative process stops and detection result is formed eventually.

$$I_t = \sum_{h}^{K}(\log(N_{C_h}^t) + \sum_{i}^{N_{C_h}^t} \frac{g(n_i)}{\sum_{j}^{N_{C_h}^t} g(n_j)} \log(\frac{g(n_i)}{\sum_{j}^{N_{C_h}^t} g(n_j)})) \tag{18}$$

$$g(n_i) = \sum_{m}^{N_{C_h}^t} P_{im} \tag{19}$$

During the tth iteration, $g(n_i)$ is the sum of shortest path distances between node n_i and internal nodes of community C_h. $N_{C_h}^t$ is the internal nodes number of community C_h. The pseudo code of iterative communities clustering is as shown in Table 2.

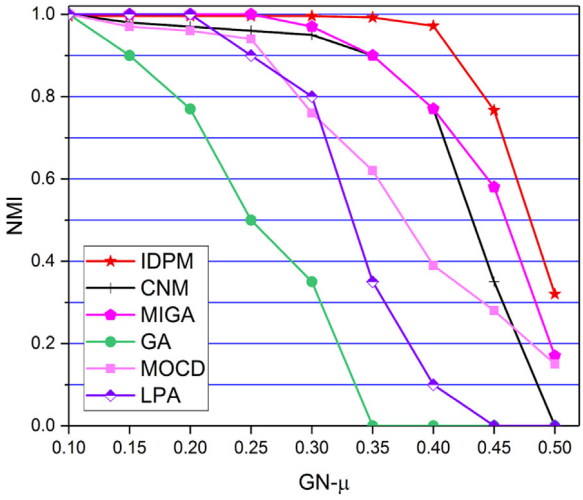
3. Experiments results and analysis

To test the performance of proposed algorithms on synthesis and real-world networks, some classical algorithms are contrasted in the experiment, including CNM [8,9], GA [11], MOGA-Net [14,29], MOCD [16], MOEAD-Net [17], Meme-Net [15], MIGA [30], MODBSA [18], LPA [19,31], CNPELPA [20] and FC [32].

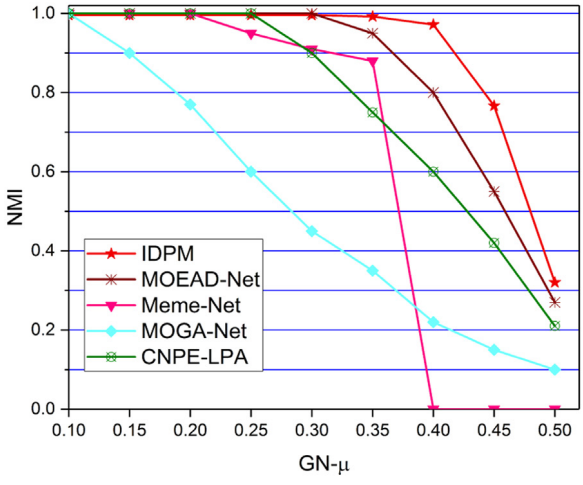
The performance of algorithms is compared by analyzing the average curve of normalized mutual information (NMI) [33] and optimal modularity [34,35].

3.1. GN synthetic network experiment

GN benchmark network is usually used to simulate the networks with small community scale, distribution of nodes in various communities uniformly and fixed degree. GN network consists of 128 nodes, which are distributed in l communities uniformly. Each community contains g nodes. The mixed parameter μ represents the probability that any node with its neighbor nodes belonging to the same community in the network.



(a)



(b)

Fig. 3. The comparison result of average NMI on GN benchmark network. The parameters $N = 128$, $l = 4$, $g = 32$, $\langle k \rangle = 16$, the range of μ is [0.1, 0.5]. (a) The average NMI result of IDPM, CNM, MIGA, GA, MOCD and LPA on GN network. (b) The average NMI result of IDPM, MOEAD, Meme, MOGA and CNPE-LPA on GN network.

In this paper, the parameters are set as follow: $l = 4$, $g = 32$, $\langle k \rangle = 16$. The mixed parameter μ varies from 0.1 to 0.5. When GN network is the status of strong connection, can be adjusted in a small range by modularity before merge of communities.

The performance of different algorithms on GN network is compared by the average NMI curve, as shown in Fig. 3. From the NMI curve observation, the following analysis can be derived:

- When mixing parameter $\mu \in [0.1, 0.35]$, the average NMI of most algorithms remain at about 1 except GA, MOGA and LPA.
- When mixing parameter $\mu \in [0.35, 0.5]$, the connecting structure of GN network have been changed immensely, which should lead to the decline of clustering accuracy. When μ is 0.45, the average NMI of IDPM is higher than others at least by 0.17. When μ is 0.5, IDPM is higher than other algorithms at least by 0.02.
- The average NMI curve of IDPM is higher than other contrasted methods on GN benchmark network.

3.2. LFR synthetic network experiment

LFR benchmark network is often used to simulate the network structure with large community scale, distribution of nodes in various communities unevenly and unfixed degree. The LFR network is closer to the actual situation of real-world network than GN network.

Table 3
The optimal modularity results of different algorithms on real network.

Algorithms	Karate	Dolphins	Football	Polbooks	Email	Power	Average
N	34	62	115	105	1133	4941	–
E	78	159	613	441	5254	6593	–
C_{max}	4	6	12	10	36	41	–
$C_{optimal}$	4	5	10	4	9	40	–
IDPM	**0.4198**	**0.5269**	**0.6046**	0.5254	**0.5733**	0.9191	**0.5947**
MOCD	0.4188	0.5259	0.5958	0.5230	0.3681	0.7065	0.5230
MOGA-Net	0.4159	0.5258	0.5280	0.5158	0.3283	0.7035	0.5028
MOEAD-Net	**0.4198**	0.5210	0.6032	0.5268	0.5193	0.6880	0.5463
GA	0.4059	0.5014	0.5940	0.5230	0.3283	0.6660	0.5031
Meme-Net	0.4020	0.5155	0.5888	0.5181	–	–	0.5061
MIGA	0.4188	0.5210	0.5911	**0.5272**	–	–	0.5145
MODBSA	**0.4198**	0.5259	**0.6046**	0.5268	0.5355	0.8140	0.5711
CNM	0.3807	0.5014	0.5766	0.5181	0.5116	**0.9345**	0.5704
LPA	0.4156	0.5268	**0.6046**	0.5262	0.5477	0.8100	0.5718
CNPELPA	0.3718	0.4833	0.6007	0.4600	–	0.8633	0.5558
FC	0.3573	0.4602	0.5793	0.4787	0.3657	0.7450	0.4977

The LFR network consists of 1000 nodes, which are distributed among several communities unevenly. The community scale is set within certain interval. The LFR network parameters γ_1 and γ_2 are degree and size distribution respectively. The variation of topology mixing parameter μ can cause the change of network connecting structure accordingly.

In this paper, LFR network parameters are set as follow: $\langle k \rangle = 20$, $k_{max} = 50$, community scale (10, 50), $\gamma_1 = 2$, $\gamma_2 = 1$, and mixing parameter μ varies from 0.1 to 0.8.

The performance of all algorithms on LFR network is compared by the average NMI curve, as shown in Fig. 4. From the NMI curve observation, the following analysis can be derived:

- When mixing parameter $\mu \in [0.1, 0.4]$, the NMI curve of most methods stay at about 0.98 1. However, the curves of CNM, MOCD, GA, MOEDA and MOGA begin to decrease partly.
- Due to parameter μ varying from 0.4 to 0.8, descend of average NMI is caused by the change of network structure. However, the NMI curve of IDPM is higher than others. When μ is 0.7, the average NMI of IDPM is higher than others at least by 0.17. When μ is 0.8, IDPM also is superior to other methods at least by 0.15.
- The average NMI of IDPM are better than other contrast algorithms in the range [0.5, 0.8], except MOEAD when is 0.8.

3.3. Real-world network experiments

In the real-world experiments, the optimal results of modularity are obtained for GN database [36], as shown in Table 3. From the comparison results, the following analysis can be derived:

- In Karate network, IDPM, MODBSA and MOEAD-Net can achieve the optimal classification result. In Football network, IDPM, LPA and MODBSA have the highest modularity, which are superior to other algorithms. In Dolphins network, the modularity of IDPM reaches the optimal result. In Email network, the IDPM's optimal modularity is better than other contrasted algorithms.
- In Polbook network, although IDPM is less than MIGA by 0.02%, MOEAD-Net and MODBSA by 0.01%, it is superior to other algorithms. In Power network, the modularity of IDPM is less than CNM by 1.5%, but it is still higher than other algorithms.
- Regarding the average optimal modularity, IDPM is higher than other contrasted algorithms by about 1% to 9%. Therefore, it can reflect that IDPM has better detection performance on real-world networks.

3.4. The analysis of parameters influence

The detection results of clustering model are affected by different parameters. S_c and N_{ckn} are the most important parameters of IDPM. More than 2000 experimental results are obtained by different parameters combinations. Modularity is usually considered about the optimal results. Therefore, the parameters influence of NMI is analyzed in this section. Parameter $\alpha = 0.1$ and $\beta = 10$. The influence of parameters about NMI is shown in Fig. 5.

From the influence diagram of parameters about NMI, the following analysis can be derived:

- In the NMI effect analysis diagram of GN network, when $\mu = 0.35$, the NMI range is [0.94, 1], and the average is 0.99. When $\mu = 0.4$, the interval of NMI is [0.9, 1], and the average is 0.98. When $\mu = 0.45$, the range of NMI changes from 0.59 to 0.93 greatly, and the interval average is 0.77. When $\mu = 0.5$, the interval of NMI is reduced, which is [0.23, 0.45], and the average is 0.31.

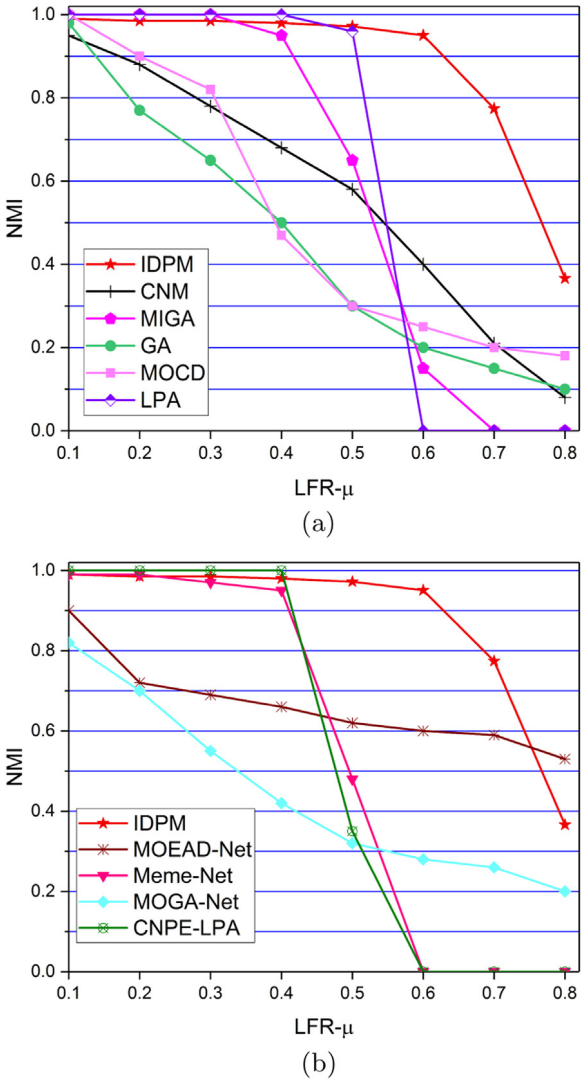


Fig. 4. The comparison result of average NMI on LFR benchmark network. The parameters $N = 1000$, $\langle k \rangle = 20$, community scale (10, 50), $k_{max} = 50$, $\gamma_1 = 2$, $\gamma_2 = 1$, the range of μ is [0.1, 0.8]. (a) The average NMI result of IDPM, CNM, MIGA, GA, MOCD and LPA on LFR network. (b) The average NMI result of IDPM, MOEAD, Meme, MOGA and CNPE-LPA on LFR network.

- In the NMI effect analysis diagram of LFR network, when $\mu = 0.5$, the NMI range is [0.96, 0.98], and the average is 0.97. When $\mu = 0.6$, the interval of NMI is [0.94, 0.96], and the average is 0.95. When $\mu = 0.7$, the range of NMI changes from 0.72 to 0.81 slightly, and the interval average is 0.77. When $\mu = 0.8$, the NMI range is [0.35, 0.38], and the average is 0.36.
- In LFR networks, the adjustment of parameters fails to cause the fluctuation of NMI. In GN networks, the variation of parameters has a slight fluctuation influence on NMI, but it can guarantee better average detection performance.

3.5. The analysis of complexity

The complexity of IDPM is mainly divided into three parts, including similarity matrix, key nodes selection and communities clustering. In the part of similarity matrix, the local complexity is $o(N^2 + N^2 \log N)$ by combining Jaccard and shortest path feature. In the part of key nodes selection, it will cost $o(2N^2)$ times by calculating ρ and δ. In the phase of communities clustering, it needs to m times to revise unstable nodes' attributes iteratively. So its complexity is regarded as $o(s(3 + m)N^2)$ where s is community number interval size. Therefore, the entire complexity of IDPM is

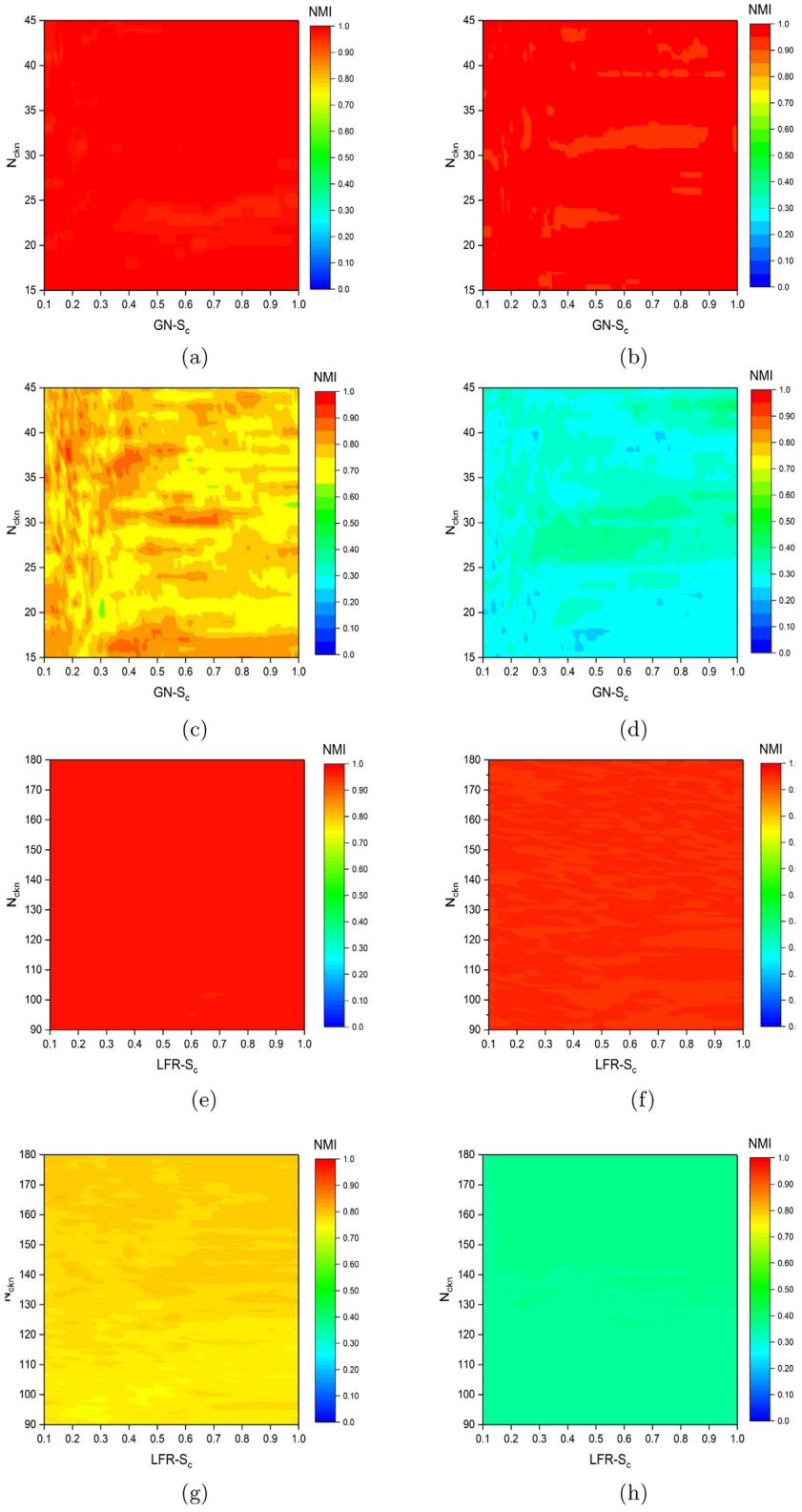


Fig. 5. The diagram of parameters influence on estimation index. (a)–(d) are the NMI effects of parameters on GN network. Parameter $\mu = 0.35$, 0.4, 0.45 and 0.5; (e)–(h) are the NMI effects of parameters on LFR-1000 network, community scale (10, 50). Parameter $\mu = 0.5$, 0.6, 0.7 and 0.8.

$o((3s+sm+3)N^2+N^2\log N)$. IDPM consumes more runtime than LPA $o(N+E)$ and CNPE-LPA $o((N+E)(E/N)^2)$. But IDPM is still faster than CNM with the complexity of $o(N^3)$.

4. Conclusion

This paper proposed a novel community detection method by improved density peak model. The composite similarity is calculated by normalizing the Jaccard and shortest path feature. The threshold conditions enables density peaks model to select representative key nodes accurately. The content of communities clustering contains the assignment of non-key nodes, merge of communities, and attributes modification of unstable nodes iteratively. The simulation results show that the detection accuracy of IDPM is superior to contrasted methods on synthetic and real-world complex networks. IDPM can be extended and transplanted to the relevant intelligent network system [37–41], network game theory [42–47] and other practical complex network applications [48–51].

Acknowledgments

This paper is supported by the National Natural Science Foundation of China (61471299), Shaanxi Province of China Key Research and Development Project (2017ZDXM-GY-139) and National Natural Science Foundation of China-Zhejiang Jointed fund (U1609216) and China CEL Network Next Generation Internet Technology Innovation Project (NGII20170606).

References

[1] R.S. Wang, S. Zhang, Y. Wang, Clustering complex networks and biological networks by nonnegative matrix factorization with various similarity measures, Neurocomputing 72 (2008) 134–141.
[2] S. Pahwa, M. Youssef, P. Schumm, Optimal intentional islanding to enhance the robustness of power grid networks, Physica A. 392 (2013) 3741–3754.
[3] F.G. Montoya, R. Baos, C. Gil, Minimization of voltage deviation and power losses in power networks using Pareto optimization methods, Eng. Appl. Artif. Intel. 23 (2010) 695–703.
[4] B. Wellman, The development of social network analysis: a study in the sociology of science by Linton C. Freeman, Soc. Net. 27 (2004) 275–282.
[5] R. Baos, C. Gil, B. Paechter, A hybrid meta-heuristic for multi-objective optimization: MOSATS, J. AppL. Math. Model. Algor. 6 (2007) 213–230.
[6] N. Wang, D. Hu, Y. Jia, Cloud computing: a literature review by using social network analysis, Inform. Resour. Manage (2015).
[7] C. Aggarwal, K. Subbian, Evolutionary network analysis: a survey, ACM. Comput. Surv. 47 (2014) 1–36.
[8] A. Clauset, M.E. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E. 70 (2004) 066111.
[9] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E. 69 (2004) 026113.
[10] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA. 99 (2002) 7821–7826.
[11] C. Shi, Z. Yan, Y. Wang, A genetic algorithm for detecting communities in large-scale complex networks, Adv. Complex. Syst. 13 (2010) 3–17.
[12] J.H. Holland, Adaptation in natural and artificial systems, Ann Arbor. 6 (1992) 126–137.
[13] M. Guerrero, F.G. Montoya, R. Baos, Adaptive community detection in complex networks using genetic algorithms, Neurocomputing 266 (2017) 101–113.
[14] C. Pizzuti, Ga-net: a genetic algorithm for community detection in social networks, in: Parallel Problem Solving from Nature-PPSN X, 10th International Conference Dortmund, 2008, p. 10811090.
[15] M. Gong, B. Fu, L. Jiao, H. Du, Memetic algorithm for community detection in networks, Phys. Rev. E. 84 (2011) 056101.
[16] C. Shi, Z. Yan, Y. Cai, Multi-objective community detection in complex networks, Appl. Soft Comput. 12 (2012) 850–859.
[17] M. Gong, L. Ma, Q. Zhang, Community detection in networks by using multi-objective evolutionary algorithm with decomposition, Physica A. 391 (2012) 40504060.
[18] F. Zou, D. Chen, S. Li, Community detection in complex networks: multi-objective discrete backtracking search optimization algorithm with decomposition, Appl. Soft Comput. 53 (2017) 285–295.
[19] U.N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, Phys. Rev. E. 76 (2007) 036106.
[20] H. Lou, S. Li, Y. Zhao, Detecting community structure using label propagation with weighted coherent neighborhood propinquity, Physica A 392 (2013) 3095–3105.
[21] M. Du, S. Ding, Y. Xue, A novel density peaks clustering algorithm for mixed data, Pattern. Recogn. Lett. 97 (2017) 46–53.
[22] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, Science. 344 (2014) 1492.
[23] T. You, H.M. Cheng, Y.Z. Ning, Community detection in complex networks using density-based clustering algorithm and manifold learning, Physica A. 464 (2016) 221–230.
[24] X. Bai, P. Yang, X. Shi, An overlapping community detection algorithm based on density peaks, Neurocomputing (2016) 226.
[25] M. Wang, W. Zuo, Y. Wang, An improved density peaks-based clustering method for social circle, Neurocomputing 179 (2016) 219–227.
[26] Q. Cheng, Z. Liu, J. Huang, Community detection in hyper network via density-ordered tree partition, Appl. Math. Comput. 276 (2016) 384–393.
[27] W. Zhang, C. Jiang, Y. Ma, An improved dijkstra algorithm based on pairing heap, in: ISCID, 2012.
[28] M. Dehmer, M. Grabner, K. Varmuza, Information indices with high discriminative power for graphs, Plos One 7 (2012) e31214.
[29] M.R. Mirsaleh, M.R. Meybodi, A michigan memetic algorithm for solving the community detection problem in complex network, Neurocomputing 214 (2016) 535–545.
[30] R. Shang, J. Bai, L. Jiao, Community detection based on modularity and an improved genetic algorithm, Physica A 392 (2013) 1215–1231.
[31] H. Zhang, X. Chen, J. Li, Fuzzy community detection via modularity guided membership- degree propagation, Pattern. Recogn. Lett. 70 (2016) 66–72.
[32] P.G. Sun, Community detection by fuzzy clustering, Physica A 419 (2015) 408–416.
[33] L. Danon, A. Diaz-Guilera, J. Duch, Comparing community structure identification, J. Stat. Mech. Theory. Exp. 2005 (2005) P09008.
[34] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.
[35] J. Liu, Fuzzy modularity and fuzzy community structure in networks, Eur. Phys. J. B. 77 (2010) 547–557.
[36] M.E.J. Newman, Network data, 2013, (http://www-personal.umich.edu/mejn/netdata/).
[37] S. Zhang, M.G. Liang, H.J. Li, Method to enhance traffic capacity for two-layer complex networks, CAN. J. Phys. 92 (2014) 1599–1605.

[38] H.J. Li, Z. Bu, Z. Wang, Enhance the performance of network computation by a tunable weighting strategy, IEEE Trans. Emerg. Top. Comput. Intell. 2 (2018) 214–223.
[39] H.J. Li, Z. Bu, Y. Li, Evolving the attribute flow for dynamical clustering in signed networks, Chaos Solitons Fractals 110 (2018) 20–27.
[40] J. Xiang, H.J. Li, Z. Bu, Critical analysis of (Quasi-) surprise for community detection in complex networks, Sci. Rep-UK. 8 (2018).
[41] Z. Bu, H.J. Li, J. Cao, Dynamic cluster formation game for attributed graph clustering, IEEE. Trans. Cybern. (2017) 1–14.
[42] Z. Bu, H. Li, J. Cao, Game theory based emotional evolution analysis for chinese online reviews, Knowl-Based. Syst. 103 (2016) 60–72.
[43] G. Gao, J. Gao, Z. Bu, A generalized game theoretic framework for mining communities in complex networks, Expert. Syst. Appl. (2017) S0957417417307406.
[44] Z. Wang, M. Jusup, R.W. Wang, Onymity promotes cooperation in social dilemma experiments, Sci. Adv. 3 (2017) 1601444.
[45] Z. Wang, S. Kokubo, M. Jusup, Universal scaling for the dilemma strength in evolutionary games, Phys. Life. Rev. 14 (2015) 1–30.
[46] Z. Wang, C.T. Bauch, S. Bhattacharyya, Statistical physics of vaccination, Phys. Rep. (2016) 664.
[47] C. Xia, S. Ding, C. Wang, Risk analysis and enhancement of cooperation yielded by the individual reputation in the spatial public goods game, IEEE. Syst. J. 11 (2016) 1516–1525.
[48] J.B. Wang, L. Wang, X. Li, Identifying spatial invasion of pandemics on meta-population networks via anatomizing arrival history, IEEE. Trans. Cybern. 46 (2016) 2782–2795.
[49] L. Wang. J.T. Wu, Characterizing the dynamics underlying global spread of epidemics, Nature Commun. 9 (2018) 218.
[50] X. Chen, C. Xia, J. Wang, A novel trust-based community detection algorithm used in social networks, Chaos Solitons Fractals 108 (2018) 57–65.
[51] C. Li, L. Wang, S. Sun, C. Xia, Identification of influential spreaders based on classified neighbors in real-world complex networks, Appl. Math. Comput. 320 (2018) 512–523.