Preventing epidemic spreading in networks by community detection and memetic algorithm

Shanfeng Wang [a], Maoguo Gong [b,*], Wenfeng Liu [b], Yue Wu [c]

[a] *School of Cyber Engineering, Xidian University, Xi'an, Shaanxi Province 710071, China*
[b] *School of Electronic Engineering, Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shaanxi Province 710071, China*
[c] *School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi Province 710071, China*

ARTICLE INFO

ABSTRACT

Targeted immunization is a commonly used strategy in preventing epidemic spreading. Traditional methods immunize targeted nodes based on specific global or local network structures instead of optimization. In this paper, we propose a novel community-based immunization strategy to select targeted immunization nodes based on optimization. The proposed algorithm consists of three steps. First, community structures are discovered by community detection algorithm. Second, possible candidates are narrowed down based on the structure properties of community. Finally, a novel memetic algorithm is designed to select immunization nodes from the candidate set. In the final step, epidemic threshold is adopted as objective function and then targeted immunization is formulated as an optimization problem. To solve this optimization problem, a novel memetic algorithm is designed. Experimental results demonstrate that the proposed algorithm outperforms some state-of-the-art immunization algorithms in optimizing epidemic threshold.

1. Introduction

Information propagation has attracted a great attention in the study of networks [1]. Information dissemination on the Internet and disease spreading in our society all belong to information propagation. These processes can be characterized by epidemic spreading models [2]. Virus spreading and rumor propagation bring serious damages to our health and our life, such that preventing information propagation is very meaningful. Detailed introduction of epidemic spreading can be found in [3].

Several models have been proposed to emulate diverse types of epidemic spreading [3], for example the susceptible infectious recovered (SIR) and susceptible infectious susceptible (SIS) dynamics. A nonlinear dynamical system (*NLDS*) approach is proposed to model epidemic spreading [4]. In *NLDS*, the authors found that there exists an epidemic threshold. Epidemic threshold $s = 1/\lambda_{1,A}$, where $\lambda_{1,A}$ is the first eigenvalue of the adjacency matrix of the graph. In this way, spreading threshold is related to network structure and preventing epidemic spreading in *NLDS* is modeled to find nodes that cause the maximal reduction in $\lambda_{1,A}$. Targeted immunization is formulated as optimizing the function of epidemic threshold. None algorithms have been proposed to optimize this function by targeted nodes immunization. Greedy

nodes immunization algorithm of optimizing this problem is with great computational complexity.

Preventing the spreading of virus or fake information is an important issue in the study of epidemic spreading. Among different algorithms on preventing epidemic spreading, immunization is the commonly used method. For example, if we aim to prevent diseases spreading between infected individuals and uninfected individuals, we can isolate the individuals who contact others frequently in the networks. In the reality, vaccination resources are always limited, such that targeted immunization strategies are of prime interest for preventing information spreading [5]. Then the problem of preventing epidemic spreading is formulated as the problem of finding targeted immunization nodes in the entire network. Based on the information used in the selection process, immunization nodes are selected in the networks at different levels [6]. At the macroscopic level, immunization nodes are selected based on nodes' centrality, for example, betweenness centrality. At the microscopic level, the most used immunization nodes are selected by degree of nodes. At the mesoscopic level, immunization nodes are selected based on the community structure of networks [7]. Under these immunization strategies, immunization nodes are selected based on the information of network structure, which are hard to find the optimal results of epidemic threshold.

In this paper, we propose a novel targeted immunization algorithm (termed as CMA-RT) by combining community structure and memetic algorithm. The proposed algorithm consists of

---

* Corresponding author.
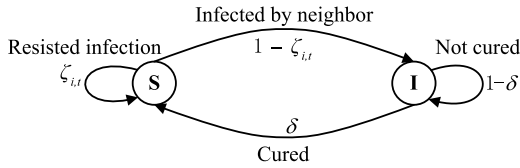 *E-mail address:* gong@ieee.org (M. Gong).

Fig. 1. An illustration of SIS model.

three steps: community detection, candidate generation and seed selection. The contributions of this paper are as follows:

- In this paper, the information of community structure, such as in-degree and out-degree of nodes, is adopted to select candidate seeds. Then immunization nodes are selected from candidate set instead of the whole network, so that it can narrow down the search space.
- In this algorithm, preventing epidemic spreading is modeled as an optimization problem by optimizing epidemic threshold. A novel memetic algorithm based on network structure is proposed to solve this problem.
- Experimental results demonstrate that the proposed algorithm outperforms some state-of-the-art global and local immunization strategies.

The remainder of this paper is organized as follows. Section 2 illustrates the related works and the motivations of the proposed algorithm. In Section 3, the proposed algorithm is introduced in detail. Section 4 gives the experimental results. Discussion is introduced in Section 5. The conclusions are summarized finally.

2. Background

2.1. Epidemic spreading models

The commonly used epidemic spreading models are the susceptible infectious recovered (SIR) and susceptible infectious susceptible (SIS) dynamics.

In [4], Chakrabarti et al. proposed a novel method to model epidemic spreading process. Fig. 1 gives the model of SIS. As Fig. 1 shows, let β be the virus birth rate on an edge connected to an infected node, and δ be virus death rate on an infected node. $p_{i,t}$ is the probability that a node i is infected at time t. $\zeta_{i,t}$ is the probability that a node i will not be infected by its neighbors in the next time-step, which is computed as follows.

$$\zeta_{i,t} = \Pi_{j \in neighbors\ of\ i}(1 - \beta * p_{j,t-1}) \qquad (1)$$

Probability $p_{i,t}$ is computed:

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{i,t} + \delta p_{i,t-1}\zeta_{i,t} \quad i = 1, \ldots, N \qquad (2)$$

Based on Eq. (2), the total number of infected nodes is

$$\eta_t = \Sigma_{i=1}^{N} p_{i,t} \qquad (3)$$

By simulations, Chakrabarti et al. found that the epidemic threshold τ of NLDS is just related to parameters β and δ. If $\beta/\delta > \tau$, the epidemic spreading will outbreak. If $\beta/\delta < \tau$, the epidemic spreading will die out over time. The epidemic threshold τ is computed as follows.

$$\tau = \frac{1}{\lambda_{1,A}} \qquad (4)$$

where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix A of the network.

In NLDS, traditional immunization strategies are no longer effective. Immunization for optimizing Eq. (4) is to find immunizations nodes to achieve maximal reduction in $\lambda_{1,A}$.

2.2. Related works

Many targeted immunization strategies have been proposed for preventing epidemic spreading.

Traditional immunization strategies select nodes based on their centralities. Degree is one of the most important properties of network [8], which is the number of neighbors of the targeted node. Computing degree of nodes is very simple and with low computational complexity, because it is related with local knowledge of network. Betweenness centrality of the targeted node is computed based on the number of shortest paths that passes through that node [9,10]. Computing betweenness centrality requires the whole knowledge of the networks. The computational complexity of betweenness centrality is high. Eigenvector centrality is computed based on the leading eigenvector of the adjacency matrix of a network [11]. Each value of the eigenvector is the score of the corresponding node. Gong et al. proposed an immunization strategy based on an expanding friendship circle [12]. Liu et al. [13] presented a local immunization strategy. In this method, the immunized node is replaced by its nonimmunized higher-score neighbor. There are also many other centralities proposed based on the global and local structures of nodes. Though these centralities are proposed for other tasks instead of epidemic spreading, these centralities can be used for selecting immunization nodes [5]. Recently, researchers have found that communities play significant roles in epidemic spreading [7,14, 15]. Many works focused on discovering immunized nodes based on the community structure [16,17]. Salathe et al. designed a strategy, the community-bridge-finder (CBF) strategy, to identify immunization targeted nodes based on random walks [7]. Gupta et al. presented a novel centrality based on community, named Comm [18]. Comm is computed on the basis of in-degree and out-degree of nodes in communities. Except for CIM, other works based on community structures adopt the information of communities, and they select the immunization nodes from the whole networks. CIM and our work narrow down the search space by community detection. Though all of these strategies above may perform well in preventing epidemic spreading, they are based on certain rules instead of optimization. They tend to obtain the near-optimal solutions instead of the optimal ones of epidemic threshold. Our work aims to optimize epidemic threshold by combining community structure and memetic algorithm.

Influence maximization problems aim to find influential nodes, such that the influence can be maximized. Influence maximization problem is similar to the problem of epidemic spreading prevention. Immunizing these influential nodes could prevent epidemic outbreaks. Influence maximization problems based on different heuristic approaches have received great attention in recent years. We proposed a discrete particle swarm optimization algorithm to optimize the local influence criterion [19]. Singh et al. proposed a learning based particle swarm optimization for influence maximization problem. Compared with other nature inspired heuristic methods for influence maximization problem, this algorithm is more effective or efficient [20]. Singh et al. presented a social spider optimization based algorithm (IM-SSO) for influence maximization problems [21] by optimizing a global influence evaluation function. They also proposed an influence maximization algorithm based on ant colony optimization (ACO-IM) by optimizing a local influence evaluation heuristic [22]. The authors extended heuristic methods to find the most influential user over multiple product diffusion multiplex networks [23]. Several efforts have done for influence maximization problems based on community structures. Wang et al. [24] proposed a community-based greedy algorithm for influence maximization problem (named as CGA). In CGA, the network is divided into different communities and influential nodes across communities. Singh et al. proposed a community based context-aware

influence maximization(C2IM) algorithm [25]. In C2IM, community detection and users' contextual features are used to select influential users. Singh et al. also proposed a community-based influence maximization (CoIM) algorithm [26]. In CoIM, the network is partitioned into several sub-networks, and then influential users are selected by local influence. Chen et al. [27] proposed a community-based strategy (named as CIM) for influence maximization problems. The proposed method consists of three steps: community detection, candidate generation and seed selection. The framework of this paper is inspired from this work. We also proposed a memetic algorithm with three steps for influence maximization problems [28]. Bozorgi et al. [29] proposed a new propagation model and exploited community structure to compute the spread of each node locally within its own community. Hajdu et al. [30] improved the efficiency of influence maximization by incorporating information on the community structure of the network into the optimization process.

2.3. Motivations of our work

As described in Introduction, immunization strategies mainly consist of global immunization strategy and local immunization strategy. As shown in the section above, it has found that bridge nodes selected based on the attributes of community detection may play more important roles in preventing epidemic spreading than those selected by global strategy. Inspired by this, community detection algorithm is adopted to divide the network into several communities in the proposed algorithm. Then, hub or bridge nodes with highest in-degree or out-degree can be selected and then candidate set of nodes are generated based on the previous obtained community structure. Based on candidate set, immunized nodes could be selected from candidate set rather than the whole network.

As shown in Section 2.1, preventing epidemic spreading can be formulated as an optimization problem by optimizing objective function Eq. (4). Optimizing objective function Eq. (4) is to find a set of nodes whose removal has a maximal decrease of Eq. (4). Suppose a network $G = (E, V)$ is a network, there are n nodes and m edges. If R nodes are immunized, the computational complexity of selecting immunized nodes is of combinatorial order $\binom{n}{R}$. Greedy algorithm can find the best solution, but high computational complexity makes it difficult to select immunized nodes in a certain time. In this way, greedy algorithm is not available for large scale networks, though it just needs to select nodes from candidate set in the proposed algorithm.

Memetic algorithm is a combination of global search and local search algorithm [31]. Global search tends to explore the promising search space efficiently, but it could just find approximate solutions. Local search finds a better solution in a local search space [32]. Memetic algorithm takes the advantages of global and local search, which balances the effectiveness and efficiency well. Memetic algorithm has been widely used in structure analysis of complex networks for its good performance in solving combinational optimization problems [33]. We proposed a memetic algorithm for epidemic spreading by structure optimization [34].

2.4. Business implications of preventing epidemic spreading

Nowadays, people communicate with others and got information conveniently by portable devices, but it may also have bad effects on social stability and personal properties. For example, computer viruses spread among routers and damage personal computers. Rumors or fake news spread on social networks rapidly, which may cause panic disorder and endangers social stability. On the contrary, if effective control strategies are developed, the damages brought by viruses, rumors and fake news may reduce to the utmost extent. Immunization strategies are commonly used by companies and governments, for example, when the computer viruses spread on the Internet, the governments and companies usually turn off the core servers or routers to prevent viruses spreading. When fake news are shared on some social networks, companies commonly limit the social media influencers to share these news so as to prevent news spreading. Then the problem above can be described as: Given a social network, how can we immunize the influencers to reduce business damages most efficiently?

Recent epidemic prevention strategies focused on global or local strategies to select immunization nodes for the maximum economic damages reduction or social stability. The proposed algorithm shows that except for influencers, immunizing individuals that link different groups also reduce economic damages in realistic and practically applications. There are also some other business implications of epidemic spread prevention, for example prevention of animal infectious diseases and urban arbovirus epidemics, etc.

It is concluded that how to prevent diseases or rumors spreading for companies or hospitals on time can reduce economic losses and save more lives.

3. The proposed algorithm

In this section, the details of the proposed algorithm will be introduced. The proposed algorithm consists of three steps. First, the network is divided into several communities by effective community detection algorithms. Second, candidate immunized nodes are generated based on the properties of community. Third, memetic algorithm is employed to select the final immunized nodes. An illustration of the proposed algorithm is shown in Fig. 2. Each step is described in detail in the following sections.
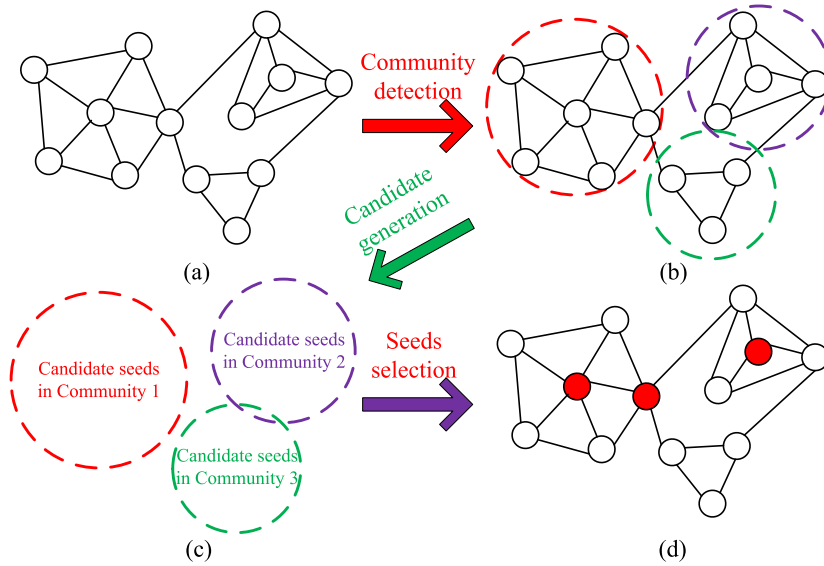
3.1. Community detection

Community is an important property of network. Community detection divides the network into several groups of nodes. Nodes in the same community are connected densely and nodes among different communities are with sparse connections. Newman et al. designed a measure, modularity (named as Q) to assess the quality of community detection [35]. Then, community detection is formulated as an optimization problem. Modularity is defined as follows.

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - k_i k_j / 2m)\delta(i, j) \tag{5}$$

where A is adjacent matrix of network, m is the number of edges in network. k_i is the degree of node i. $\delta(i, j) = 1$, if node i and j are in the same community, otherwise, 0.

A large number of community detection algorithms have been proposed by optimizing modularity [36]. Some representative community detection algorithms by maximizing modularity include heuristics algorithms [37,38] and evolutionary algorithms [39,40]. In our algorithm, the community detection algorithm in [37], named as BGLL, is used. Compared with other community detection algorithms, BGLL is with several advantages. First, BGLL performs well in discovering communities of networks. Second, the computational complexity of BGLL is very low, so that BGLL could be used for community detection in large scale networks. Last but not least, it does not need to give the number of communities in advance in BGLL. These advantages are the motivations of why BGLL is adopted for community detection. BGLL consists of two processes. First, node i is assigned to one of its neighbors communities which could generate the maximum and positive gain of modularity. Second, a new network is generated based on the community division of step 1. Then, the same process of step 1 is performed on the new network.

Fig. 2. Framework of the proposed algorithm. (a) The structure of network. (b) The results of community detection. (c) Generate the candidate nodes of each community based on community structure. (d) Select the immunized nodes based on memetic algorithm.

3.2. Candidate generation

It is still difficult to select the immunized nodes based on community detection. Though many measures have been proposed to select nodes based on the properties of community, they are not suitable for *NLDS*. Candidate generation aims to find some candidate seeds based on community, which can narrow the search space. As we discuss above, except for hub nodes, bridge nodes also play important roles in epidemic spreading, especially in networks with strong community structure. In the proposed algorithm, in-degree and out-degree of nodes in community are used to select candidate seeds.

It is a challenge in candidate generation to determine the number of immunized nodes for each community. Larger communities, in which there are more nodes, have more important effects on epidemic spreading of networks [27]. Nodes immunized in large communities will be more than those in small communities. In other word, nodes to be immunized in the proposed algorithm is proportional to the size of community. Hub nodes are usually nodes with highest centralities. In the proposed algorithm, we select top-20% nodes with highest in-degree as candidate hub nodes in each community. For bridge nodes, we select nodes whose out-degree is more than 1.

3.3. Memetic algorithm for seed selection

With candidate seeds, it is still difficult to select the final immunized nodes. Greedy algorithm for selecting immunized nodes from candidate seeds is with large computational complexity. To reduce computational complexity, a novel memetic algorithm is designed to select immunized nodes from candidate seeds. In this section, the proposed memetic algorithm for seed selection is described in Algorithm 1. The immunized nodes are selected by optimizing function (4).

Here, we give an explanation of different components in Algorithm 1. Initialization is used to generate an initial population based on candidate seeds. Selection is used to select individuals of high quality or representative individuals from parent population. GeneticOperation aims to generate offsprings by crossover or mutation operators and LocalSearch aims to search better individuals. Update aims to update new population based on parent population and newly generated population.

Algorithm 1 Framework of the proposed memetic algorithm for seed selection

Input: Adjacency matrix of network: *A*, candidate seeds: *Candidate*, number of removed nodes: *R*, maximum number of generations: *maxgen*, population size: *pop*, mating pool size: *pool*, tournament size: *tour*, crossover probability: P_c, mutation probability: P_m.

Output: Immunized nodes.

$\mathbf{P} \leftarrow$ Initialization(*pop*)

For $i = 1$ to *maxgen* **do**

1. $\mathbf{P}_{parent} \leftarrow$ Selection($\mathbf{P}$, *pool*, *tour*)
2. $\mathbf{P}_{child} \leftarrow$ GeneticOperation($\mathbf{P}_{parent}$, P_c, P_m)
3. $\mathbf{P}_{new} \leftarrow$ LocalSearch($\mathbf{P}_{child}$)
4. $\mathbf{P} \leftarrow$ Update($\mathbf{P}$, $\mathbf{P}_{new}$)

End for

3.3.1. Representation and initialization

The proposed algorithm aims to select a set of nodes to immunize. In the proposed memetic algorithm, each chromosome represents an individual or a set of immunized nodes. Each chromosome is encoded as an integer string.

$$x = \{x_1, x_2, \ldots, x_R\} \tag{6}$$

where R is the number of immunized nodes. x_i is the ith node that need to be immunized selected from candidate seeds. In each chromosome, each node is just immunized one time.

For each chromosome, every element or gene is generated randomly from candidate. To accelerate the convergence of the proposed algorithm, the first chromosome is replaced by the list which has the maximum degree, because the degree strategy is demonstrated with good performance.

3.3.2. Genetic operators

In memetic algorithm, genetic operators are used to generate new chromosomes, such as crossover and mutation. Genetic operators can prevent the algorithm being trapped into local optimum. In the proposed algorithm, crossover and mutation are designed based on the knowledge of network and communities.
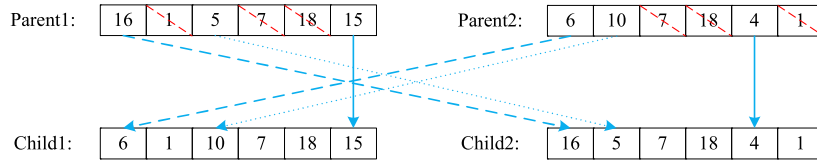
Fig. 3. Illustration of crossover operator. In this figure, 6 nodes are immunized. There are common 3 nodes in both of parent chromosomes and these nodes will remain unchanged. Nodes 16, 5, 15 in parent chromosome 1 and nodes 6, 10, 4 will be exchanged. Suppose random numbers of these three pairs of nodes be 0.7, 0.6 and 0.3. Node 16 and 6 will be swapped. Node 5 and 10 are also swapped. Node 4 and 15 will keep unchanged.
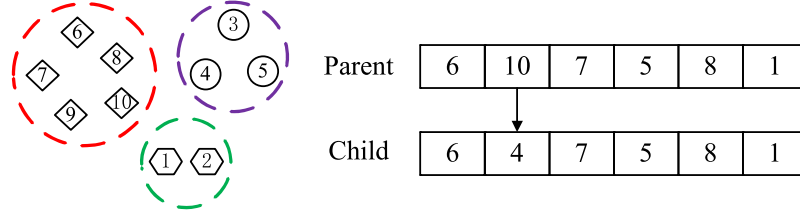
Fig. 4. Illustration of mutation operator. In this figure, there are three communities. Nodes 1 and 2 are in community 1. Nodes 3, 4 and 5 are in community 2. Nodes 6, 7, 8, 9 and 10 are in community 3. Node 10 will perform mutation, and then one node selected from community 1 or 2 replaces node 10.

Crossover. In this paper, a random crossover is adopted. Given two chromosomes A and B, the common elements in A and B keeps unchanged. For elements in just one chromosome, a random number is generated and if the random number is more than 0.5, and then corresponding elements will be swapped. Fig. 3 shows the process of crossover. Offspring chromosomes generated by this crossover operator could inherit most of elements in parent chromosomes.

Mutation. One point mutation in the proposed algorithm is adopted. For each node in one chromosome x_A, a random number is generated. If the random number is larger than mutation probability p_m, the selected node v_i is replaced by node v_j, where v_j does not belong to chromosome x_A and v_j and v_i are in different communities. An illustration of the mutation process is shown in Fig. 4.

3.3.3. Local search

In memetic algorithm, genetic operators help to explore global search scope, which is easily trapped into local optimum. To prevent the proposed algorithm being trapped into local optimum, local search is used to improve solution in local search scope. In the proposed crossover, offspring chromosomes are generated by exchanging parts of two parent chromosomes. None new nodes can be found. Mutation operator focuses on exploring nodes in other communities. In the proposed memetic algorithm, a neighbor-based local search is presented.

Suppose node i be in community C, neighbors of node i are all the nodes in the community C except for node i. In the proposed local search, each node i is replaced by one of the nodes in the neighbors of node i, which could get the best optimum. Local search in the proposed algorithm is just performed on the chromosome with the best fitness in the population. Local search is illustrated in Algorithm 2. The proposed local search focuses on exploring space within community.

In Algorithm 2, $Eval()$ is to compute the fitness of the chromosome. The FindBestNeighbor() function aims to find the neighbor chromosome with the largest fitness value.

4. Experimental results

In this section, experiments are performed on several real world networks to validate the performance of the proposed algorithm in preventing the outbreaks of epidemic spreading. All

---

Algorithm 2 The local search procedure

Input: chromosome: X_{parent}, number of removed nodes: R
Output: chromosome: X_{child}.

$X_{current} \leftarrow X_{parent}$;
For each node $i = 1$ to R **do**

    1. $X_{next} \leftarrow$ FindBestNeighbor($X_{current}$)
    2. **if** $Eval(X_{next}) > Eval(X_{current})$
    3. $X_{current} \leftarrow X_{next}$;
    4. **end if**

End for $X_{child} \leftarrow X_{current}$;

---

the experiments are performed on a machine with Intel Core i7-8700 CPU (3.2 GHZ), 64 GB RAM, running Windows 10.[1]

4.1. Datasets

In this paper, we use six real-world networks to show the performance of the proposed algorithm. Properties of these networks are shown in Table 1.

Dolphin. Dolphin network was introduced by Lusseau, after he observed behaviors of 62 Bottlenose dolphins for 7 years [41]. In this network, node represents dolphin and edge in this network represents dolphin's significant frequent association.

Sfi. Sfi network is a scientific collaboration network [42]. In this network, node represents scientists and edges represents collaborations among scientists.

Elegans. Elegans network is a metabolic network [43]. In this network, node represents organism and edge represents the interaction between them.

Email. Email network is an email network of University Spain, containing 1669 users in different industries [44]. In this network, each node is an email and edge represents communication of each pair of nodes.

Hamsterster. Hamsterster network is a network of the friendships and family links between users of the website [45]. In this network, nodes represent users and edges denote their relationships between users.

---

[1] The source code for CMA-RT can be found at https://web.xidian.edu.cn/sfwang/index.html.

Table 1
Properties of networks used in this algorithm.

Networks	The number of nodes	The number of edges
Dolphins	62	159
Sfi	198	2742
Elegans	453	2025
Email	1133	5451
Hamsterster	2426	16 630
Ca-Grqc	5242	14 490

Table 2
The values of parameters used in this algorithm.

Parameters	The meaning of parameter	Values of parameters
pop	The size of population	100
maxgen	The number of generations	100
pool	Size of the mating pool	100
tour	Tournament size	2
p_c	The probability of crossover	0.9
p_m	The probability of mutation	0.1

Table 3
Community detection results in this algorithm.

Networks	Modularity	The number of communities
Dolphin	0.5188	5
Sfi	0.7497	8
Elegans	0.4266	10
Email	0.5406	11
Hamsterster	0.5473	177
Ga-Grqc	0.8010	395

Ca-Grqc. Ca-Grqc is a collaboration network of Arxiv General Relativity category [46]. In this network, each node represents an author and each edge represents that two authors coauthor papers.

4.2. Experimental settings and comparison algorithms

To validate the performance of the proposed algorithm, some targeted immunization strategies, including global and local immunized strategies, are adopted as comparison algorithms.

Degree. Degree is a strategy of selecting immunized nodes by degree of nodes. The top-*R* nodes with highest degree are immunized. In this strategy, degree of each node is computed and then top-*R* nodes with the highest degree are immunized. This strategy is a global immunization strategy.

Betweenness. Betweenness is a strategy of immunizing top-*R* nodes with the highest betweenness centrality. In node betweenness strategy, first each node betweenness is computed and then top-*R* nodes with highest betweenness are removed from the network. This strategy is also a global immunization strategy.

Comm. Comm is the algorithm introduced in [18]. Immunized nodes in Comm are selected by in-degree and out-degree of nodes in communities. In this strategy, network is divided into several communities and then comm of each node is computed. Top-*R* nodes are removed from the network based on comm of each node. This strategy is a local immunization strategy based on community.

Adaptive degree. Adaptive degree is the centrality where the degree is re-calculated after the removal of each node. In this section, adaptive degree is named as Adade.

Adaptive betweenness. Adaptive betweenness is the centrality where the betweenness is re-calculated after the removal of each node. In this section, adaptive betweenness is named as Adabet.

CIM. This method is proposed for influence maximization problem based on community detection [27]. This method also consists of three steps. In the experiments, the nodes which have the largest influence are immunized.

LAPSO. This method is a learning based particle swarm optimization for influence maximization problem. Compared with other nature inspired heuristic methods for influence maximization problem, LAPSO is more effective or efficient [20].

In the proposed algorithm, there are also some parameters. These parameters are set by trail and error to make the performance of the proposed algorithm perform well on different networks. Parameters used in the proposed algorithm are shown in Table 2.

4.3. Experimental results

In the proposed algorithm, the first step is community detection. As described in Section 2.3, there are mainly two goals of using community detection. The first one is adopting the properties of communities to generate candidate seeds. By candidate seeds, it could just select target immunized nodes from candidate seeds, in such way, the search space reduces. Table 3 gives the results of community detection, including the values of modularity and the number of communities. By selecting nodes based on the community structure of networks, the number of candidate seeds is much smaller than the number of nodes in the networks, so that the search space is narrowed down.

In the following experiments, the results of different algorithms on six networks are given. These results aim to validate the ability of the proposed algorithm to optimize the epidemic threshold Eq. (4). When immunizing different number of nodes, results of these algorithms on six networks are shown in Fig. 5. In the experiments, average results of Comm, LAPSO and the proposed algorithm over 10 times are adopted for comparison. In Fig. 5, 5%, 10% and 15% nodes of each network will be immunized.

On dolphins network, betweenness-based strategy and adaptive betweenness strategy are the worst performers. Degree-based strategy, LAPSO, CIM and adaptive degree strategies also have good performances, but they are worse than the proposed algorithms. On Sfi network, CIM and betweenness-based strategies have worst performance. Adaptive degree-based strategy and LAPSO have similar results. When more than 5% nodes are immunized, Comm method could not find effective immunization nodes. The proposed algorithm outperforms other algorithms. On the elegans network, except the CIM algorithm, other algorithms all have good abilities to prevent epidemic spreading. When 15% nodes are immunized, adaptive degree method gets the best result. On the Email network, the proposed algorithm, adaptive degree, degree-based and LAPSO all have great performance, while CIM, Comm, betweenness and adaptive betweenness methods perform worse. On the hamsterster network, CIM and betweenness are difficult to discover immunization nodes. On the Ca-Grqc network, except the proposed algorithm, degree and adaptive degree strategies, other methods performed not very well.

It is concluded that the proposed CMA-RT almost outperforms other algorithms in optimizing the objective function Eq. (4) on six networks. Degree-based and adaptive degree-based strategies also perform well on six networks. LAPSO performs well on the former five networks but it performs worse at Ca-Grqc network. Comm, betweenness and adaptive betweenness algorithms do not perform well on most networks.

The reasons why the proposed algorithm performs well are as follows. First, the proposed algorithm gets immunization nodes by optimizing Eq. (4) directly while other strategies do not. Second, in CMA-RT, candidate seeds are selected based on community and in the selection of candidate nodes, nodes are selected based on in-degree and out-degree of nodes. Most nodes with
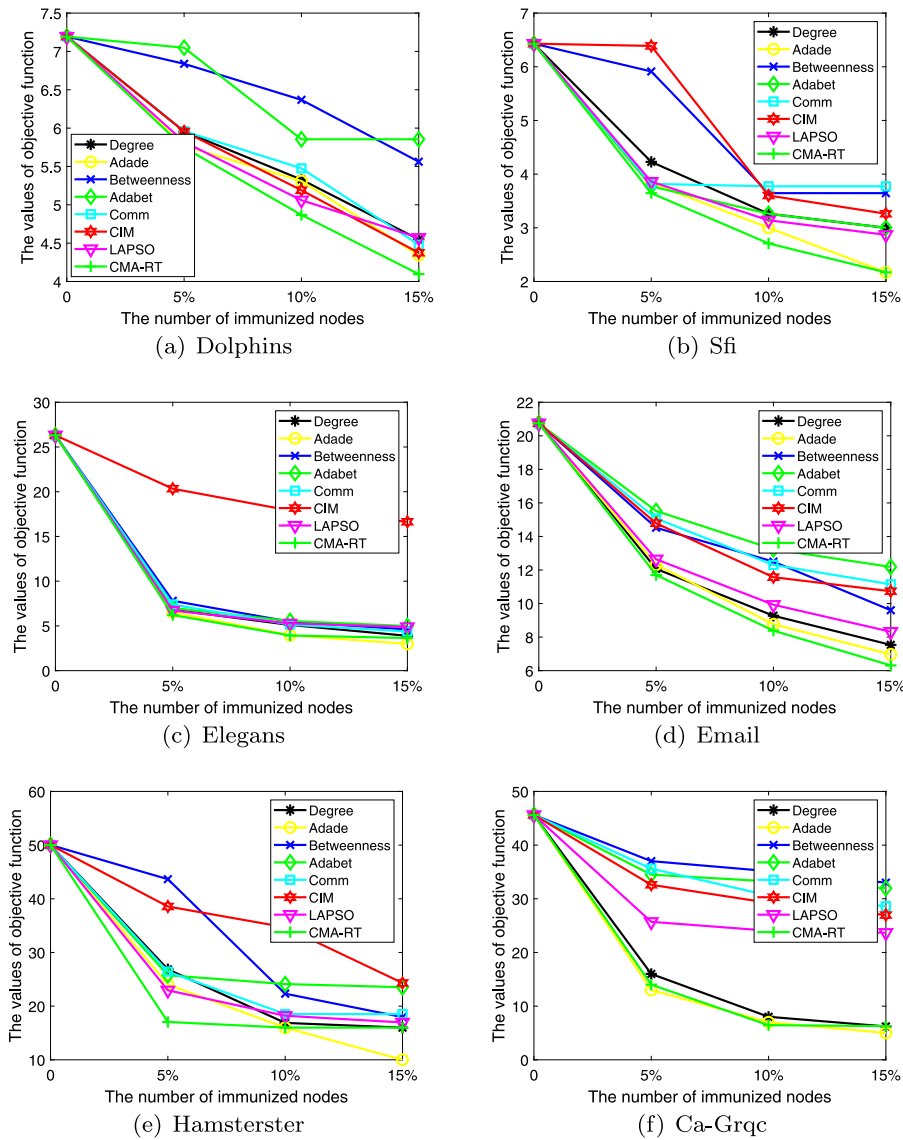
Fig. 5. The values of objective function obtained by different algorithms on real networks.

the highest degree are included in candidate nodes. Compared with degree-based strategy, the proposed algorithm also considers bridge nodes. Third, the novel proposed memetic algorithm has a good performance of selecting final immunized nodes from candidate seeds.

To show the robustness of the proposed algorithm, the proposed algorithm is run 10 times on these networks. The boxplots are given in Fig. 6. As Fig. 6 shows, on dolphins, Sfi and hamsterter networks, our algorithms can achieve similar results when it is run 10 times. On other networks, though the results are different each time, they are similar and acceptable. Fig. 6 shows that the proposed algorithm has robust results because the proposed algorithm is heuristic and it combines the advantages of global and local searches.

To detect the significant differences in the epidemic spreading prevention, the Friedman test [47] is used to test whether there are significant differences. In this part, the level confidence α is set 0.05. The Friedman test results are given in Table 4 and the results indicate that there are significant difference in preventing epidemic spreading. And then the Holm–Bonferroni procedure is adopted as a post-doc procedure to detect the concrete differences between algorithms. In the test, the proposed

Table 4
The Friedman test on epidemic spreading prevention.

#Nodes	p-values	Hypothesis
5%	1.77E4	Rejected
10%	2.36E5	Rejected
15%	6.74E5	Rejected

algorithm is selected as the control algorithm. Table 5 shows that the proposed algorithm performs statistically better than other algorithms in preventing epidemic spreading.

As shown in Fig. 5, degree-based strategy is worse than the proposed algorithm. In the following experiments, to illustrate the competitive advantages of CMA-RT over other algorithms clearly, only degree-based strategy is used to make a comparison and just experiments on the smallest dolphins network are shown. Fig. 7 shows the results of two different algorithms on dolphins networks, when different number of nodes are immunized. As Fig. 7 shows, the proposed algorithm always has better results than degree-based strategy when different number of nodes are immunized. Suppose $\beta = 0.1$ and $\delta = 0.6$, for degree-based immunization strategy, epidemic spreading will die
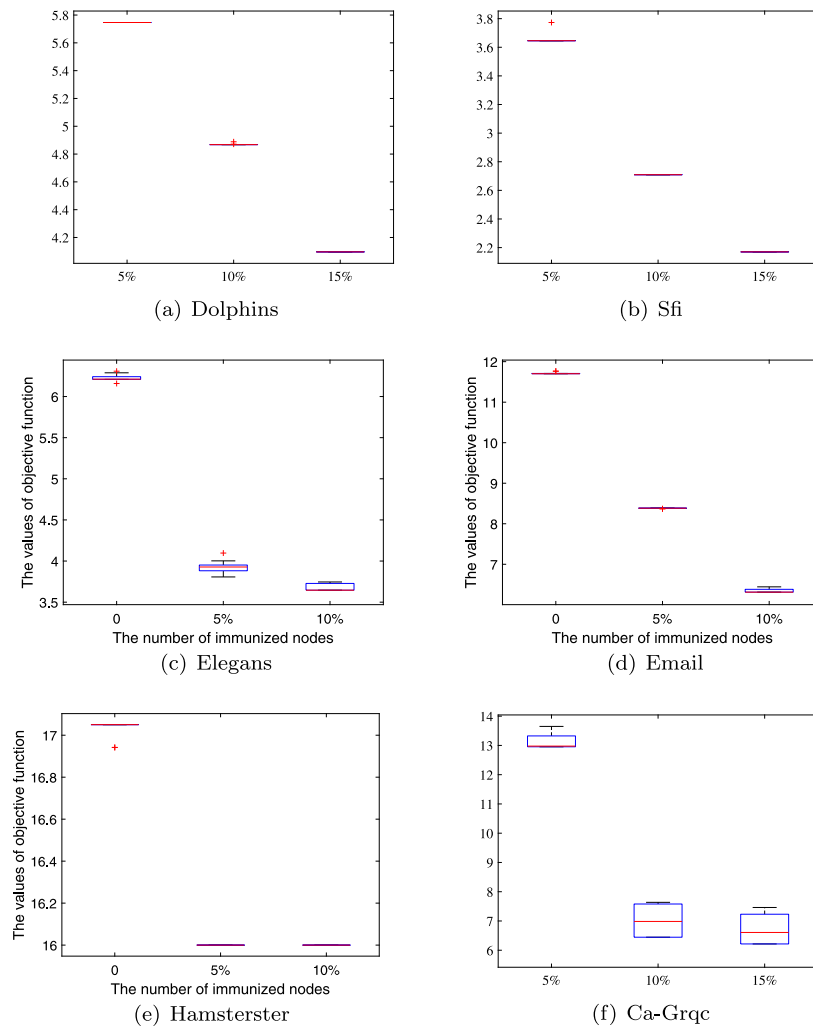
(a) Dolphins

(b) Sfi

(c) Elegans

(d) Email

(e) Hamsterster

(f) Ca-Grqc

Fig. 6. The statistical results of the proposed algorithm on real networks.

Table 5

The p-value/i results based on the Holm–Bonferroni procedure on the experimental results.

#Nodes	Degree	Adade	Betweenness	Adabet	Comm	CIM	LAPSO	CMA-RT
5%	0.0143/1	0.1025/7	0.0143/2	0.0143/3	0.0143/4	0.0143/5	0.0143/6	-/*
10%	0.0143/1	0.0253/6	0.0143/2	0.0143/3	0.0143/4	0.0143/5	0.0253/7	-/*
15%	0.1797/6	0.6547/7	0.0143/1	0.0143/2	0.0143/3	0.0143/4	0.0143/5	-/*

out when 3 nodes are immunized. For the proposed algorithm, epidemic spreading will die out when 2 nodes are immunized. Suppose $\beta = 0.2$ and $\delta = 1.1$, 6 and 4 nodes are needed to be immunized for degree-based strategy and CMA-RT, respectively. It is concluded that to prevent the spreading of epidemic, the proposed algorithm needs fewer nodes to be immunized. In reality, this result is very important when vaccination resources are limited.

In Fig. 8, the network structure is given. As shown in Fig. 8, dolphins network is divided into 5 communities. Here, we take dolphins network to the reasons why degree-based strategy is not better than the proposed algorithm. Suppose 3 nodes are immunized, degree-based strategy will immunize nodes 15, 38 and 46, while the proposed algorithm CMA-RT will immunize nodes 2, 15 and 46. The degree of node 38 is 11 and the degree of node 2 is 8. Node 2 is also not the one with the highest degree in it community. From the community division of network, we can find that node 2 connects two communities and its out-degree is 2. In other word, node 2 is a bridge node and its immunization plays an important role in preventing epidemic spreading.

Fig. 9 shows the number of infected nodes when different number of nodes are immunized on dolphins network. When more nodes are immunized, the threshold will increase, such that epidemic is hard to outbreak. When β/δ is less than threshold, the epidemic will die out. When β/δ is larger than threshold, the epidemic will outbreak. With the increase of β/δ, more nodes will be infected.
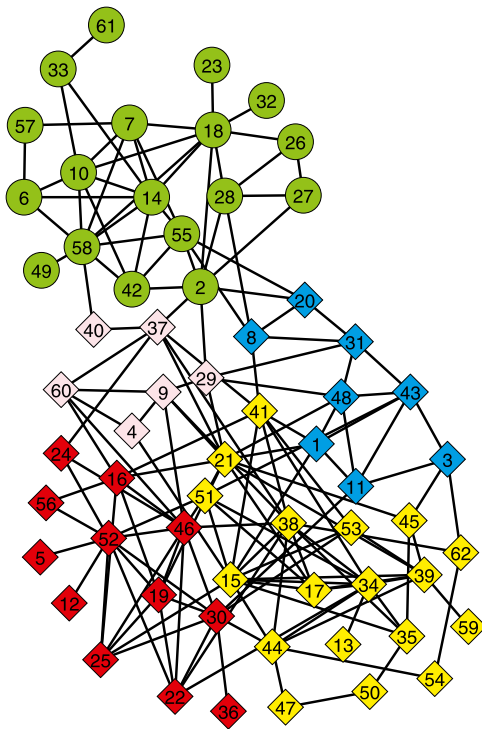
4.4. Computational complexity

In this section, the computational complexity of the proposed algorithm is given. Given a network with n nodes and m edges. The proposed algorithm is divided into three processes. The computational complexity of community detection algorithm BGLL is $O(n \log n)$. Suppose n_c nodes are selected as candidate seeds and R nodes are immunized. In the process of candidate generation, the computational complexity is $O(n^2)$. In the process of seed generation, the computational complexity of memetic algorithm is computed as follows. In each iteration of memetic
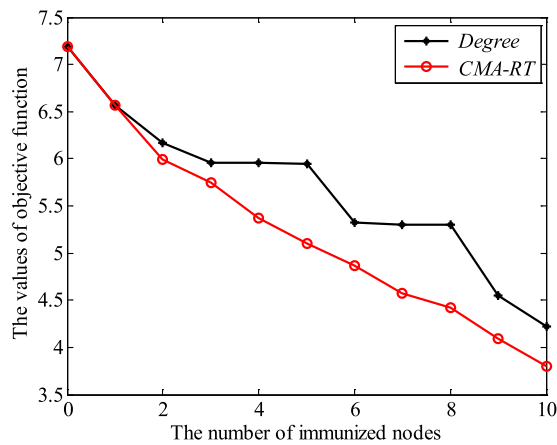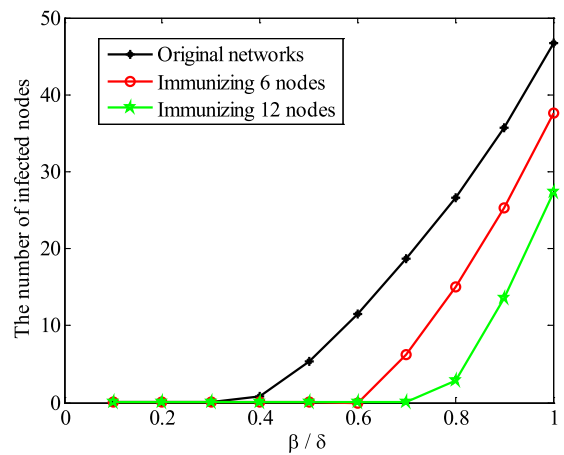
Fig. 7. Comparison of degree-based strategy and the proposed CMA-RT on dolphins network.

Fig. 9. The number of infected nodes in dolphins network when different number of nodes are immunized by the proposed CMA-RT.

Fig. 8. An illustration of dolphins networks. BGLL algorithm divides this network into 5 communities.

algorithm, crossover and mutation are performed $[pool/2]$ times and $[pool]$ times. In this paper, the largest eigenvalue is computed by power iteration and the computational complexity is $O(n^2)$. Then, the computational complexity of genetic operators is $O(pool(R + n^2))$. In the local search, it is needed to consider $\log R$ neighbors for each node. The computational complexity of local search is $O(R \log R n^3)$. The computational complexity of Update is $O(pop + pool)$. Then the computational complexity of the proposed algorithm is $O(R \log R n^2)$. In other word, the computational complexity of the proposed algorithm is high because the computational complexity of computing threshold function is high. In memetic algorithm, all the chromosomes in each population are independent of each other. To reduce the computational time, the objective functions of chromosomes in the population are computed parallel in this paper.

5. Discussion

Chakrabarti et al. found that epidemic threshold is related to the largest eigenvalue of the adjacency matrix A of the network. Then, epidemic spreading prevention is formulated as an optimization problem by optimizing epidemic threshold in the proposed algorithm. To deal with this problem, an algorithm based on community detection and memetic algorithm is proposed.

Though experimental results validate that the proposed algorithm outperforms state-of-the-art algorithms in optimizing epidemic threshold, there are still some limitations of the proposed algorithm. First, the computational complexity of the proposed algorithm is still high. Second, the number of immunization nodes should be determined in advance, and the proposed algorithm could not find the minimized number of immunization nodes to prevent epidemic spreading outbreaking.

6. Conclusions

This paper proposed a novel framework to prevent epidemic spreading by optimization. The proposed algorithm combines community structure information and memetic algorithm to select targeted immunization nodes. Candidate seeds are narrowed down based on the community structure information. A memetic algorithm is proposed to find the final immunization nodes by optimizing epidemic threshold from candidate seeds. Compared with greedy algorithm, memetic algorithm is with lower computational complexity and good performance. Experiments on six networks showed that the proposed algorithm outperforms traditional immunization strategies in achieving maximal reduction of largest eigenvalue of the adjacency matrix.

The proposed algorithm provides a novel method based on optimization instead of targeted immunization strategies related to network structures. In the future, we will focus on targeted immunization in large scale of networks by designing more effective candidate generation strategy and memetic algorithm.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106118.

CRediT authorship contribution statement

Shanfeng Wang: Conceptualization, Methodology, Software, Funding acquisition, Writing - original draft, Writing - review & editing. **Maoguo Gong:** Conceptualization, Methodology, Funding acquisition, Writing - review & editing. **Wenfeng Liu:** Visualization, Investigation, Writing - review & editing. **Yue Wu:** Supervision, Validation, Writing - review & editing.

Acknowledgments

References

[1] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (14) (2001) 3200.
[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (4–5) (2006) 175–308.
[3] R. Pastorsatorras, C. Castellano, P.V. Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Modern Phys. 87 (3) (2014) 120–131.
[4] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos, Epidemic thresholds in real networks, ACM Trans. Inf. Syst. 10 (4) (2008) 1.
[5] S. Wandelt, X. Sun, D. Feng, M. Zanin, S. Havlin, A comparative analysis of approaches to network-dismantling, Sci. Rep. 8 (1) (2018) 13513.
[6] L. Hébertdufresne, A. Allard, J.G. Young, L.J. Dubé, Global efficiency of local immunization on complex networks, Quant. Biol. 7 (8) (2013) 5295–5301.
[7] M. Salathé, J.H. Jones, Dynamics and control of diseases in networks with community structure, PLoS Comput. Biol. 6 (4) (2010) e1000736.
[8] R. Pastor-Satorras, A. Vespignani, Immunization of complex networks, Phys. Rev. E 65 (3) (2002) 036104.
[9] Z. Hai-Feng, L. Ke-Zan, F. Xin-Chu, W. Bing-Hong, An efficient control strategy of epidemic spreading on scale-free networks, Chin. Phys. Lett. 26 (6) (2009) 068901.
[10] M.E. Newman, A measure of betweenness centrality based on random walks, Soc. Netw. 27 (1) (2005) 39–54.
[11] Y. Yamada, T. Yoshida, A comparative study of community structure based node scores for network immunization, in: Int. Conf. Active Media Tech., Springer, 2012, pp. 328–337.
[12] K. Gong, M. Tang, P.M. Hui, H.F. Zhang, D. Younghae, Y.-C. Lai, An efficient immunization strategy for community networks, PLoS One 8 (12) (2013) e83489.
[13] Y. Liu, Y. Deng, B. Wei, Local immunization strategy based on the scores of nodes, Chaos 26 (1) (2016) 013106.
[14] W. Huang, C. Li, Epidemic spreading in scale-free networks with community structure, J. Stat. Mech.-Theory E 2007 (01) (2007) P01014.
[15] X. Wu, Z. Liu, How community structure influences epidemic spread in social networks, Physica A 387 (2) (2008) 623–630.
[16] N. Masuda, Immunization of networks with community structure, New J. Phys. 11 (12) (2009) 123018.
[17] H. Yang, M. Tang, H.-F. Zhang, Efficient community-based control strategies in adaptive networks, New J. Phys. 14 (12) (2012) 123017.
[18] N. Gupta, A. Singh, H. Cherifi, Centrality measures for networks with community structure, Physica A 452 (2016) 46–59.
[19] M. Gong, J. Yan, B. Shen, L. Ma, Q. Cai, Influence maximization in social networks based on discrete particle swarm optimization, Inform. Sci. 367 (2016) 600–614.
[20] S.S. Singh, A. Kumar, K. Singh, B. Biswas, LAPSO-IM: A learning-based influence maximization approach for social networks, Appl. Soft Comput. (2019) 105554.
[21] S.S. Singh, A. Kumar, K. Singh, B. Biswas, IM-SSO: Maximizing influence in social networks using social spider optimization, Concurr. Comput.: Pract. Exper. e5421.
[22] S.S. Singh, K. Singh, A. Kumar, B. Biswas, ACO-IM: maximizing influence in social networks using ant colony optimization, Soft Comput. (2019) 1–23.
[23] S.S. Singh, K. Singh, A. Kumar, B. Biswas, MIM2: Multiple influence maximization across multiple social networks, Physica A 526 (2019) 120902.
[24] Y. Wang, G. Cong, G. Song, K. Xie, Community-based greedy algorithm for mining top-k influential nodes in mobile social networks, in: Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, ACM, 2010, pp. 1039–1048.
[25] S.S. Singh, A. Kumar, K. Singh, B. Biswas, C2IM: Community based context-aware influence maximization in social networks, Physica A 514 (2019) 796–818.
[26] S.S. Singh, K. Singh, A. Kumar, B. Biswas, Coim: Community-based influence maximization in social networks, in: Int. Conf. Adv. Informat. Comput. Research, Springer, 2018, pp. 440–453.
[27] Y.-C. Chen, W.-Y. Zhu, W.-C. Peng, W.-C. Lee, S.-Y. Lee, CIM: community-based influence maximization in social networks, ACM Trans. Intell. Syst. Tech. 5 (2) (2014) 25.
[28] M. Gong, C. Song, C. Duan, L. Ma, B. Shen, An efficient memetic algorithm for influence maximization in social networks, IEEE Comput. Intell. Mag. 11 (3) (2016) 22–33.
[29] A. Bozorgi, S. Samet, J. Kwisthout, T. Wareham, Community-based influence maximization in social networks under a competitive linear threshold model, Knowl.-Based Syst.
[30] L. Hajdu, M. Krász, A. Bóta, Community based influence maximization in the Independent Cascade Model, in: 2018 Federated Conf. Comput. Sci. Infor. Syst., FedCSIS, IEEE, 2018, pp. 237–243.
[31] Y.-S. Ong, M.H. Lim, X. Chen, Research frontier-memetic computation—past, present & future, IEEE Comput. Intell. Mag. 5 (2) (2010) 24.
[32] R. García-Ródenas, L. Jiménez Linares, J. Alberto López-Gómez, A memetic chaotic gravitational search algorithm for unconstrained global optimization problems, Appl. Soft Comput. 79 (2019) 14–29.
[33] L. Ma, M. Gong, J. Liu, Q. Cai, L. Jiao, Multi-level learning based memetic algorithm for community detection, Appl. Soft Comput. 19 (2014) 121–133.
[34] J. Yan, M. Gong, L. Ma, S. Wang, B. Shen, Structure optimization based on memetic algorithm for adjusting epidemic threshold on complex networks, Appl. Soft Comput. 49 (2016) 224–237.
[35] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2) (2004) 026113.
[36] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (3) (2010) 75–174.
[37] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech.-Theory E 2008 (10) (2008) P10008.
[38] M. Girvan, M.E. Newman, Community structure in social and biological networks, Proc. Nat. Acad. Sci. USA 99 (12) (2002) 7821–7826.
[39] Q. Cai, M. Gong, L. Ma, S. Ruan, F. Yuan, L. Jiao, Greedy discrete particle swarm optimization for large-scale social network clustering, Inform. Sci. 316 (2015) 503–516.
[40] F. Cheng, T. Cui, Y. Su, Y. Niu, X. Zhang, A local information based multi-objective evolutionary algorithm for community detection in complex networks, Appl. Soft Comput. 69 (2018) 357–367.
[41] D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, S.M. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, Behav. Ecol. Sociobiol. 54 (4) (2003) 396–405.
[42] M. Gong, Q. Cai, L. Ma, S. Wang, Y. Lei, Computational Intelligence for Network Structure Analytics, Springer, 2017.
[43] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E 72 (2) (2005) 027104.
[44] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, A. Arenas, Self-similar community structure in a network of human interactions, Phys. Rev. E 68 (6) (2003) 065103.
[45] R. Rossi, N. Ahmed, The network data repository with interactive graph analytics and visualization, in: AAAI, Vol. 15, 2015, pp. 4292–4293.
[46] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data 1 (1) (2007) 2.
[47] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.