An Overlapping Community Detection Based Multi-Objective Evolutionary Algorithm for Diversified Social Influence Maximization

Lei Zhang†, Fengjiao Sun†, Fan Cheng†, Haiping Ma‡ and Xiaoyan Sun*
†*Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Computer Science and Technology, Anhui University, Hefei 230039, China*
Email: zl@ahu.edu.cn, summerfyannis@gmail.com, chengfan@mail.ustc.edu.cn (*corresponding author*)
‡*Institutes of Physical Science and Information Technology, Anhui University, Hefei 230039, China*
Email: hpma2020@163.com
**School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China*
Email: xysun78@126.com

***Abstract*—Influence maximization refers to selecting a group of nodes from a social network, which obtains the largest influence spread under a cascade model. However, most of the existing works only focused on the influence and ignored the diversity of influenced crowd. Thus, scholars have raised the issue of diversified social influence maximization recently, using the category information of nodes to design diversity indicator and introducing a trade-off parameter to balance the two objectives *influence* and *diversity* as one single objective for optimization. In fact, the category information of nodes in the network is usually difficult to be collected, thus the definition of diversity based on nodes' categories is not very general and accurate. In addition, it is very difficult to set the trade-off parameter, especially when there is no prior knowledge in real applications. To this end, we employ overlapping community structure information to design the diversity of nodes without any node's additional (e.g. category) information. Due to the two objectives of *influence* and *diversity* may be conflicting, a multi-objective evolutionary algorithm named MOEA-DIM is proposed to optimize the two objectives simultaneously, which does not need to set the trade-off parameter between the two objectives. In MOEA-DIM, a network reduction strategy based on overlapping community structure is suggested to greatly reduce the search space. In addition, a population initialization strategy based on random walk is designed to accelerate the convergence of the algorithm. Experiments on six real-world datasets show that the proposed algorithm MOEA-DIM has promising performance in terms of both effectiveness and efficiency.**

***Index Terms*—Influence maximization, diversity, overlapping community detection, multi-objective evolutionary algorithm.**

I. Introduction

Today, online social networks such as Facebook and Twitter play a great role in people's life. People share their opinions through social platform to make information easier to spread between users [1]–[3]. The problem of influence maximization is to simply find top-k high-influence users in social networks, which has the maximum influence. Kempe *et al.* [4] formalized the problem as a discrete optimization problem and proved that the problem was an NP-hard problem and then used a greedy algorithm to solve the problem. The greedy algorithm has a certain degree of accuracy, but it used Monte Carlo simulation to calculate the marginal effects of all nodes to select nodes, which has a high time complexity and cannot be used in large-scale networks. Based on this issue, scholars have proposed various greedy [5]–[7] and heuristic strategies [8]–[14] to accelerate speed of the algorithm.

However, most of the existing studies only focused on the influence of the influenced crowd and ignored the importance of diversity. In fact, in real-world marketing, enterprises prefer to choose diverse target audience to promote new products and expand the market size. And in investing, investors are more inclined to choose diverse portfolios to reduce investment risks. To this end, Tang *et al.* [15] first utilized the category information of the nodes to design *diversity* indicator. Then they formulated a novel problem of diversified social influence maximization by using a trade-off parameter to combine the two objectives *influence* and *diversity* as one single objective for optimization. At the same time, they proved that this problem was an NP-hard problem and then proposed a greedy algorithm to solve it. Finally, the experimental results on two real datasets validated the effectiveness of their proposed problem and algorithm. However, in real world, the category information of nodes is usually difficult to be collected. Thus, designing diversity indicator based on this additional information is not very accurate and universal to a certain extent. In addition, the trade-off parameter between *influence* and *diversity* is difficult to set, especially when there is no prior knowledge of the application can be considered.

To solve the above challenges, we use the overlapping community detection algorithm to obtain community structure information of nodes and employ these information to design novel diversity indicator without using any additional (e.g. category) information of nodes. In the real world, overlapping community detection is an important tool for discovering information hidden in complex networks such as social networks, people with similar hobbies and interests are more likely to be divided into same community and one people maybe have

978-1-7281-6929-3/20/$31.00 ©2020 IEEE

more than one communities or hobbies. Thus, it is intuitive to use the community structure information of nodes to design diversity indicator. In addition, we propose a multi-objective evolutionary algorithm to simultaneously optimize *influence* and *diversity* without setting the trade-off parameter between the two objectives. In summary, the major contributions of this paper are summarized as follows:

- We employ the overlapping community structure information of nodes to design novel diversity indicator, which does not need any node's additional information. Considered that the two objectives of *influence* and *diversity* are conflicting, we formulate the diversified social influence maximization as a multi-objective optimization problem.
- We propose a multi-objective evolutionary algorithm named MOEA-DIM to solve the above multi-objective optimization problem. In MOEA-DIM, in order to reduce the search space, we suggest a network reduction strategy by using overlapping community structure information to prune out unimportant nodes. In addition, we design a population initialization strategy based random walk to accelerate the convergence of the algorithm.
- We estimate the effectiveness and efficiency of MOEA-DIM on six real-world networks with different characteristics. Experimental results show that the proposed algorithm MOEA-DIM has better performance than the comparison algorithms.

The remainder of this paper is organized as follows. We first give the preliminaries and related work in Section II. Then, we describe the problem model and introduce the proposed MOEA-DIM algorithm in Section III. Section IV shows our experimental results and Section V concludes this paper.

II. The Preliminaries and Related Work

In this section, we first present some preliminaries about influence maximization and traditional diversified social influence maximization problem, then introduce the multi-objective optimization. Finally, we give the related work.

A. Influence Maximization Problem

For the first time, Kempe *et al.* [4] formalized the social influence maximization problem as a discrete optimization problem. Formally, given a graph $G = (V, E)$, the influence maximization problem is defined as follows:

$$\max_{S \subseteq V} \sigma(S), \quad s.t. \ |S| = k, \tag{1}$$

where $\sigma(S)$ denotes the influence of activated nodes by the seed set S and k represents the size of S. Kempe *et al.* proved that this problem was an NP-hard one, then they used Monte Carlo simulations to calculate the influence of influenced crowd and proposed a (1-1/e) approximation greedy algorithm.

B. Traditional Diversified Social Influence Maximization Problem

Tang *et al.* [15] proposed the problem of diversified social influence maximization. In their work, they used the category information of the nodes in the network to measure the diversity of nodes. For example, in movie recommendation dataset, nodes represent movies and the category of the node is the category of the movie. If the type of one movie is comedy, love and action at the same time, then this node has three categories. They assumed that each node belongs to one or more categories, which is described as a distribution over these categories. Based on these distributions, they designed a novel diversity measure $D(\mu^S)$, which has the nice property of monotonicity and submodularity. Then, they combined the two goals of diversity and influence into one objective to optimize. The diversified social influence maximization problem is defined as follows:

$$\max \ F(S) = (1-\gamma)\frac{\sigma(S)}{\bar{\sigma}} + \gamma\frac{D(\mu^S)}{\bar{D}}, \quad s.t. \ |S| = k. \tag{2}$$

where $\bar{\sigma}$ and $\bar{D}$ are two normalization factors, γ is a trade-off parameter. They proved that this objective function was non-decreasing submodular, and then employed a simple greedy algorithm to address it.

C. Multi-Objective Optimization

Multi-objective optimization problem (MOP) refers to the optimization of multiple objectives at the same time, and multiple objectives usually conflict with each other. In real applications, a variety of tasks can be formulated as MOPs, such as pattern recommendation [16], [17], community detection [18], [19], network vulnerability analysis [20], and so on.

A multi-objective optimization problem (take the maximization problem as an example) can be formally defined as follows:

$$max \quad \mathbf{F}(x) = (f_1(x), f_2(x), ..., f_m(x))^T, \tag{3}$$

where $x = (x_1, x_2, ..., x_n) \in \Omega$ is the decision vector, Ω is the n-dimensional decision space and m is the number of objectives.

Suppose two decision vectors x_1 and x_2, x_2 dominates x_1 or x_1 is dominated by x_2 (expressed as $x_2 \succ x_1$) if $f_i(x_1) \leq f_i(x_2)$ for all $i = 1, 2, ..., m$, and $\mathbf{F}_i(x_1) \neq \mathbf{F}_i(x_2)$. The non-dominated solution $x \in \Omega$ is a decision vector, if there is no any $x^* \in \Omega$ satisfying $x^* \succ x$. The set of all non-dominated solutions is represented as a Pareto set defined as $PaS = \{x \in \Omega | \nexists x^* \in \Omega, x^* \succ x\}$. The Pareto front is a projection of the Pareto set to the target space, which is denoted as $PF = \{\mathbf{F}(x) | x \in PaS\}$. The purpose of the multi-objective evolutionary algorithm is to search a group of non-dominated solutions that approximate the true Pareto front.

D. Related Works

Kempe *et al.* [4] formalized the social influence maximization problem into a discrete optimization problem for the first time and proved that it is an NP-hard problem. Then they proposed a (1-1/e) approximation greedy algorithm to solve it. However, the proposed greedy algorithm is not very efficient and it is difficult to be applied for large-scale networks. For this reason, a large number of scholars have proposed various greedy and heuristic strategies to accelerate speed of the algorithm [5]–[14].

To be specific, on the one hand, some scholars proposed to reduce the number of calculations of the influence of seed to improve the efficiency. For example, Leskovec *et al.* [5] used the submodule property to propose a CELF optimization strategy to reduce the user's search space. Based on the similar idea, Goyal *et al.* [6] proposed a CELF++ optimization strategy, which can further reduce the number of calculations of seed influence. Chen *et al.* [7] proposed the NewGreedy and MixGreedy algorithms to simulate the influence propagation process by generating a subgraph.

On the other hand, some scholars have focused on heuristic algorithms to improve the efficiency. Heuristic algorithms usually select seed nodes according to some heuristic rules, whose advantage is that the speed is fast, but there is no guarantee of accuracy. For example, Chen *et al.* [9] proposed a PMIA model by using the local influence of the nodes to approximate the global influence and constructing the maximum influence subtree through the Maximum Influence Path (MIP). Gong *et al.* [10] presented an efficient memetic algorithm based on community to reduce the network and adopted the 2-hop influence spread [21] method to reduce computational cost. Recently, Huang *et al.* [14] proposed a new method integrated community detection into the process of influence propagation by digging potential topic information and community members to infer the correlation between users and finding influential nodes about communities.

Most of the above existing works have been devoted to designing efficient algorithms to solve the problem of maximizing influence, but a few researchers focused on the diversity of influenced people. For example, Tang *et al.* [15] first proposed the problem of diversified social influence maximization, where the influence and the diversity of influenced crowd were both considered. They used the category information of nodes to design the diversity indicator, and the two objectives of influence and diversity were combined as one objective by setting a trade-off parameter for optimization. In fact, the category information of nodes is usually difficult to be collected in the real world. In addition, the trade-off parameter is difficult to set for different networks with different characteristics. Caliò *et al.* [22] defined a novel problem, named Diversity-sensitive Targeted Influence Maximization (DTIM) and brought the concept of topology-driven diversity into targeted influence maximization problems. They assumed that users' diversity in a social graph can be determined based on topological properties related to their neighbors. Finally, they employed a greedy approach that exploited the search for shortest paths in the diffusion graph, in a backward fashion from the selected target set.

Different from the above related works, this paper proposes to employ overlapping community structure information to design the diversity of nodes without any node's additional information. Due to the conflict between influence and diversity, we propose a multi-objective algorithm to simultaneously optimize influence and diversity. In MOEA-DIM, two novel strategies (i.e. a network reduction strategy based on overlapping community structure information and population initialization strategy based on random walk) are proposed to greatly reduce the search space and accelerate convergence of the algorithm.

III. The Proposed Algorithm for Diversified Influence Maximization

In this section, we propose a novel diversity indicator based on overlapping communities and then formalize the diversified social influence maximization as a multi-objective optimization problem. In what follows, we introduce the proposed multi-objective evolutionary algorithm named MOEA-DIM, including the proposed network reduction strategy and population initialization strategy in detail.

A. The Overlapping Community Based Diversity and Multi-Objective Problem Formulation

The overlapping community detection algorithm is firstly used to obtain community structure information of nodes, which can be used to design the diversity indicator. To be specific, each community is assigned with one label, then the labels of communities where one node is belonged to are used as this nodes's label information. It can be found that each node belongs to one or more labels, described as a distribution over these labels. Then, the popular Shannon entropy is adopted to calculate the diversity of influenced crowd by using nodes' distribution over these labels. To be specific, the proposed diversity indicator named $StructDiversity$ is formulated as follows.

Definition 1 (StructDiversity): Given a seed S, the struct diversity for the network is defined as:

$$SD\left(\mu^S\right) = \sum_{i=1}^{|C|} -p_i \log_2 p_i,$$
$$where \quad p_i = \frac{\sum\limits_{j=1}^{|V|} w_{ij}\mu_j^S}{\sum\limits_{i=1}^{|C|}\sum\limits_{j=1}^{|V|} w_{ij}\mu_j^S} \tag{4}$$

where μ^S is a vector with length $|V|$ that represents the influence of nodes in the network activated by seed S, $C = \{c_1, c_2, ..., c_m\}$ is a division of G with m communities, p_i can be understood as the proportion of influence allocated to the i-th community, μ_j^S is the probability of node j being activated after Monte Carlo simulation, w_{ij} is the label distribution of node j in the community i. For each node j in the network, a vector of length $|C|$ satisfies $\sum_{i=1}^{|C|} w_{ij} = 1$. It can be found that the proposed diversity indicator just uses the inner struct information of nodes without use any additional (e.g. category) information of nodes, thus is more general and accurate.

Let's take the network shown in Fig. 1 as an example. Suppose this network is divided into two communities (labeled with A and B respectively), where blue nodes 5 and 7 are two overlapping nodes. Let seed S be $\{1, 5\}$. After 10,000 times Monte Carlo simulations, suppose the probability of each node being activated in 10,000 times are $\mu_1^S = 1.0$, $\mu_4^S = 0.3$, $\mu_5^S = 1$, $\mu_7^S = 0.1$, $\mu_8^S = 0.2$ respectively, and

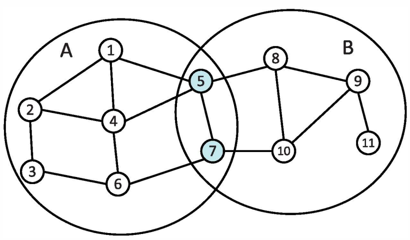


Fig. 1. A network is divided into two communities A and B, where blue nodes $\{5, 7\}$ are overlapping nodes.

for other nodes, their probabilities are all 0. We can find that the labels of nodes {1,4,5,7,8} are $L_1 = \{A\}$, $L_4 = \{A\}$, $L_5 = \{A, B\}$, $L_7 = \{A, B\}$, $L_8 = \{B\}$ respectively. Therefore, the label distribution of node 1 in the community A, $w_{A1} = 1$. Similarly, $w_{A4} = 1$, $w_{A5} = w_{B5} = 0.5$, $w_{A7} = w_{B7} = 0.5$, $w_{B8} = 1$. The proportion of influence allocated to the A, B community (no normalization) is $p'_A = 1 * 1 + 0.3 * 1 + 1 * 0.5 + 0.1 * 0.5 = 1.85$, $p'_B = 1*0.5+0.1*0.5+0.2*1 = 0.75$. After normalization, $p_A = 1.85/(1.85+0.75) = 0.71$, $p_B = 0.75/(1.85+0.75) = 0.29$. Then $SD(\mu^S) = -(p_A \log_2 p_A + p_B \log_2 p_B) = 0.52 + 0.36 = 0.87$.

Tang *et al.* [15] have verified the conflict between influence and category-based diversity through their experiments. In other words, when the diversity is large, the influence may be small. Otherwise, when the diversity is small, the influence may be large. This observation is also verified in our experiments shown in Section IV-B. To this end, it is very intuitive to formalize the diversified social influence maximization as a multi-objective optimization problem by optimizing both influence $\sigma(S)$ and struct diversity $SD(\mu^S)$ of influenced crowd, which is defined as follows:

$$Maximize \quad \mathbf{F}(S) = (f_1(S), f_2(S))^T,$$

$$where \quad \begin{cases} f_1(S) = \sigma(S) \\ f_2(S) = SD\left(\mu^S\right) \end{cases} \tag{5}$$

In the following, an effective multi-objective evolutionary algorithm MOEA-DIM is proposed for this multi-objective optimization problem.

B. The Framework of MOEA-DIM

In this paper, we propose a multi-objective evolutionary algorithm for diversified social influence maximization, which adopts a similar framework with NSGA-II [23] to solve the multi-objective problem. The integer encoding is adopted as the individual encoding. To be specific, the i-th individual in the population includes a k-node set, which is denoted as $p_i = \{x_1, x_2, ..., x_k\}$ and x_j is one node in the network.

Algorithm 1 shows the general framework of MOEA-DIM, which consists of four steps: overlapping community detection, network reduction, population initialization and population evolution. At the first step, we use one popular overlapping community detection algorithm SLPA [24] to get the overlapping community structure C. At the second step, in order to reduce the search space, a network reduction strategy based on overlapping community structure information is proposed for pruning out unimportant nodes to obtain the candidate node set G' (see Algorithm 2). At the third step, a population initialization strategy based random walk is suggested to accelerate the convergence of the proposed algorithm (see Algorithm 3). At the last step, MOEA-DIM uses tournament selection to obtain mating pool and employs parents in the mating pool to crossover and mutation to produce child individuals as suggested in [23], where each node in the child individuals should be appeared in G'.

From the above explanation of Algorithm 1, it can be found that the proposed network reduction strategy and population initialization strategy are two important parts in MOEA-DIM. In the following, we will illustrate them in detail.

Algorithm 1: General Framework of MOEA-DIM

Input: G: the network of $G = (V, E)$; $maxgen$: the maximum number of generations; pop: the size of population;
Output: Optimal solutions OS

Step1: Overlapping Community Detection
$C \leftarrow$ obtain community structure and node label information by a popular overlapping community detection algorithm SLPA [24];
Step2: Network Reduction
$G' \leftarrow$ obtain node candidates by **the proposed Algorithm 2**;
Step3: Population Initialization
$P_0 \leftarrow$ obtain initial population by **the proposed Algorithm 3**;
$t = 0$;
Step4: Population Evolution
while $t < maxgen$ **do**
  $P_t \leftarrow Evaluate(P_t)$;
  $M_t \leftarrow GenerateMatingPool(P_t)$;
  $Q_t \leftarrow CrossMutate(M_t, G')$;
  $P_{t+1} \leftarrow UpdatePopulation(P_t \cup Q_t)$;
  t++;
end while
$OS \leftarrow$ get the non-dominated solutions from the final population P_t;

C. The Proposed Strategies in MOEA-DIM

In this section, we will introduce the two proposed strategies including network reduction strategy and population initialization strategy.

1) Network reduction strategy: The basic idea of network reduction is that some unimportant nodes will increase the search space of the algorithm, so we harness some useful information to prune out some unimportant nodes in the network. To be specific, we design two indicators (denoted as $StructDegree$ and $NeiDiversity$) by considering two aspects of influence and diversity to select node candidates.

Note that the overlapping node in the network is like bridge connecting different communities and plays a very important role in network information diffusion [25]. To this end, we propose a novel structural degree indicator by using nodes' structural information of the overlapping nodes to measure the influence importance of nodes in the network. Given a network $G = (V, E)$, let n be the number of all nodes in V, L_v represents the community labels of the node v and

N_v represents the neighbors of node v. The structural degree indicator can be defined as follows.

Definition 2 (StructDegree):

$$StructDegree\,(v) = \frac{|L_v|+1}{2} + \sum_{u \in N_v} \frac{|L_u|+1}{2}, \tag{6}$$

where $(|L_v|+1)/2$ represents the overlapping importance of node v itself, $\sum\limits_{u \in N_v} (|L_v|+1)/2$ represents the overlapping importance of node v's neighbors. From the formula, we can see that if a node is not an overlapping node, then its contribution to structural degree is 1. This means that the more overlapping nodes a node connects, the greater structural degree of the node.

For example, as shown in Fig. 1, nodes 4, 5 have four neighbors with the same degree. The neighbors of node 4 and 5 are $\{1, 2, 5, 6\}$ and $\{1, 4, 7, 8\}$ respectively, According to Eq. (6), $StructDegree(4) = 1 + (1 + 1 + 1.5 + 1) = 5.5$ and $StructDegree(5) = 1.5 + (1 + 1 + 1.5 + 1) = 6$. We can find that node 5 is better than node 4. Although node 4 and node 5 have same degree, node 5 is more important since it is an overlapping node. From this example, we can see that the importance between overlapping nodes and non-overlapping nodes with the same degree can be distinguished according to the proposed indicator.

In addition, it is intuitive to find that if a node's neighbors belong to more different communities, then the diversity of n-odes will be greater according to the Definition 1. Therefore, a new indicator named neighbor diversity of node v is proposed to measure the diversity importance of nodes in the network, which is defined as follows:

Definition 3 (NeiDiversity):

$$NeiDiversity\,(v) = \sum_{i=1}^{|C|} f_i \left(\sum_{u \in N_v} w_{iu} \right) \tag{7}$$

where $f_i(x)$ can be any nondecreasing concave function, we set $f_i(x) = ln(1+x)$ and N_v represents the neighbors of node v, w_{iu} is the distribution of node u in the community i.

Also taking the network shown in Fig. 1 as an example. The network is divided into two communities labeled as A and B respectively, where blue nodes 5, 7 are two overlapping nodes. Take node 5 as an example. The neighbors of node 5 are $\{1, 4, 7, 8\}$, and the community labels of node 1, 4, 7, 8 are $L_1 = \{A\}$, $L_4 = \{A\}$, $L_7 = \{A, B\}$ and $L_8 = \{B\}$ respectively. Therefore, the nodes in A community are $\{1, 4, 7\}$ and the nodes in B community are $\{7, 8\}$. So, the neighbour diversity in community A is $f_A = ln(1 + 1 + 1 + 0.5) = 1.25$, and the neighbour diversity in community B is $f_B = ln(1 + 0.5 + 1) = 0.92$. Finally, the neighbour diversity for node 5 is $NeiDiversity(v_5) = f_A + f_B = 2.17$.

Algorithm 2 presents the procedure of network reduction strategy, which uses the two suggested indicators $StructDegree$ and $NeiDiversity$ to prune out some u-nimportant nodes in the network. First, we calculate the $StructDegree$ and $NeiDiversity$ value of each node. Then, we sort the nodes in descending order according to $StructDegree$ and $NeiDiversity$ value respectively. To automatically get a proper number of candidates, we select $\sqrt{nk}$ nodes as candidates for each indicator, where n is the number of nodes in the network and k is the size of seed S. Then, we merge and remove duplicate nodes for the two candidates as one final candidates, whose length is much smaller than n. Therefore, the search space can be greatly reduced by using this strategy.

Algorithm 2: Network Reduction

Input: G: the network of $G = (V, E)$; n: the number of nodes in network G; k: seed size;
Output: CN: Candidate Nodes

$CN \leftarrow \emptyset$;
for each $u \in V$ **do**
  Calculate $StructDegree(u)$ by Eq. (6);
  Calculate $NeiDiversity(u)$ by Eq. (7);
end for
$SD \leftarrow$ select top-$\sqrt{nk}$ nodes with high $StructDegree$;
$ND \leftarrow$ select top-$\sqrt{nk}$ nodes with high $NeiDiversity$;
$CN \leftarrow SD \cup ND$;

Algorithm 3: Population Initialization

Input: *pop*: population size; k: seed size;
Output: Population P

$P = \{p_1, p_2, ..., p_{pop}\} \leftarrow \emptyset$;
$p_1, p_2 \leftarrow$ use a greedy algorithm to respectively obtain one most influential k nodes and one most diverse k nodes;
for $i = 3$ to *pop* **do**
  $t = 0$;
  while $t < k$ **do**
    $(v, u) \leftarrow$ randomly select two different nodes from candidate nodes CN obtained by Algorithm 2;
    if $RScore(v) > RScore(u)$ and $v \notin p_i$ **then**
      $p_i \leftarrow p_i \cup \{v\}$;
      $t = t + 1$;
    end if
    if $RScore(v) \leq RScore(u)$ and $u \notin p_i$ **then**
      $p_i \leftarrow p_i \cup \{u\}$;
      $t = t + 1$;
    end if
  end while
end for

2) Population initialization strategy: In multi-objective evolutionary algorithms, a random strategy is usually used to generate the initial population. However, randomly initialized individual solutions usually deteriorate the algorithm's con-vergence in diversified social influence maximization. To this end, we propose a population initialization strategy based on random walk to accelerate the convergence of the proposed algorithm. Algorithm 3 presents the procedure of population initialization strategy. First, for the two objectives influence and diversity, we use a greedy algorithm to obtain two good individuals (p_1 and p_2) with k nodes, which has the largest value of influence or diversity. Then, for the remaining indi-viduals, we select better nodes from the candidates to generate individuals by using the binary tournament mechanism.

Since random walk is often used to calculate the importance of nodes in a network, we use a popular indicator of random walk score $RScore$ [26] to measure the importance of nodes

TABLE I
REAL-WORLD NETWORKS WITH DIFFERENT CHARACTERISTICS.

Networks	#Nodes	#Edges	Average Degree	Maximal Degree
Erdos	6.9k	11.8k	3.42	507
HepTh	9.9k	26.0k	5.26	65
Anybeat	12.6k	67.1k	5.30	4,516
AstroPh	18.8k	198.1k	21.10	504
Email-enron	33.7k	180.8k	10.73	1,383
Gemsec-RO	41.8k	125.8k	6.02	112

in binary tournament mechanism, which is defined as follows.

$$RScore\,(u) = \sum_{v \in N_u} \frac{preRScore\,(v)}{d\,(v)}, \tag{8}$$

where $preRScore$ is the random walk score at the previous iteration, $d(v)$ represents the degree of node v.

Finally, base on Equation (8), *pop* individuals can be generated by using the binary tournament mechanism.

IV. Experimental Results

In this section, we first give experimental setup, including datasets and comparison algorithms. Then, we provide sufficient experimental results to verify the effectiveness and efficiency of the proposed algorithm MOEA-DIM compared with the baseline algorithms. Finally, we show the effectiveness of the two strategies proposed in MOEA-DIM.

A. Experimental Setup

1) Datasets: To evaluate the performance of proposed algorithm, six real-world networks with different characteristics are adopted. Among them, Anybeat network [27] and Gemsec-RO network [27] are social networks. Erdos network [28], HepTh network [29], and AstroPh network [29] are collaboration networks. Email-enron network [27] is a communication network. The basic characteristics of these six real-world networks are given in Table I, where #Nodes and #Edges denote the number of nodes and edges in the network respectively.

2) Comparison Algorithms: We compare MOEA-DIM with baseline algorithms and its variants.

- D-Inf: D-Inf is a greedy algorithm by using a trade-off parameter γ to balance two objectives *influence* and *diversity* as one objective for optimization [15].
- MOEA-DIM(-Red): In order to verify the effectiveness of the proposed network reduction strategy, we also compare MOEA-DIM with a variant named MOEA-DIM(-Red), where we set all parts the same as MOEA-DIM except there is no network reduction strategy.
- MOEA-DIM(-Ini): In order to verify the effectiveness of the proposed population initialization strategy, we also compare MOEA-DIM with a variant named MOEA-DIM(-Ini). MOEA-DIM(-Ini) keeps the same as MOEA-DIM except the random initialization strategy is used.
- MOEA-DIM(EDV): In order to reduce the calculation time of MOEA-DIM, the approximate method EDV [30] is used in MOEA-DIM to calculate the influence by replacement of Monte Carlo simulation. And this variant is named as MOEA-DIM(EDV).

TABLE II
PARAMETERS SETTINGS OF COMPARISON ALGORITHMS.

Parameter	Meaning	Value
$maxgen$	the maximum generation	100
pop	population size	100
$pool$	the size of mating pool	100
pm	mutation probability	1/k
$iters$	the maximum iteration of SLPA	100
θ	the threshold of SLPA	0.1
t	the times of random walk	3
pp	propagation probability in IC model	0.005

TABLE III
EFFICIENCY COMPARISON (IN MINUTES).

Networks	D-Inf	MOEA-DIM	MOEA-DIM(EDV)
Erdos	16.69	8.11	0.27
HepTh	16.16	2.80	0.14
Anybeat	2,150.37	38.72	1.88
AstroPh	231.04	29.59	1.27
Email-enron	424.96	52.60	2.72
Gemsec-RO	79.72	4.51	0.51

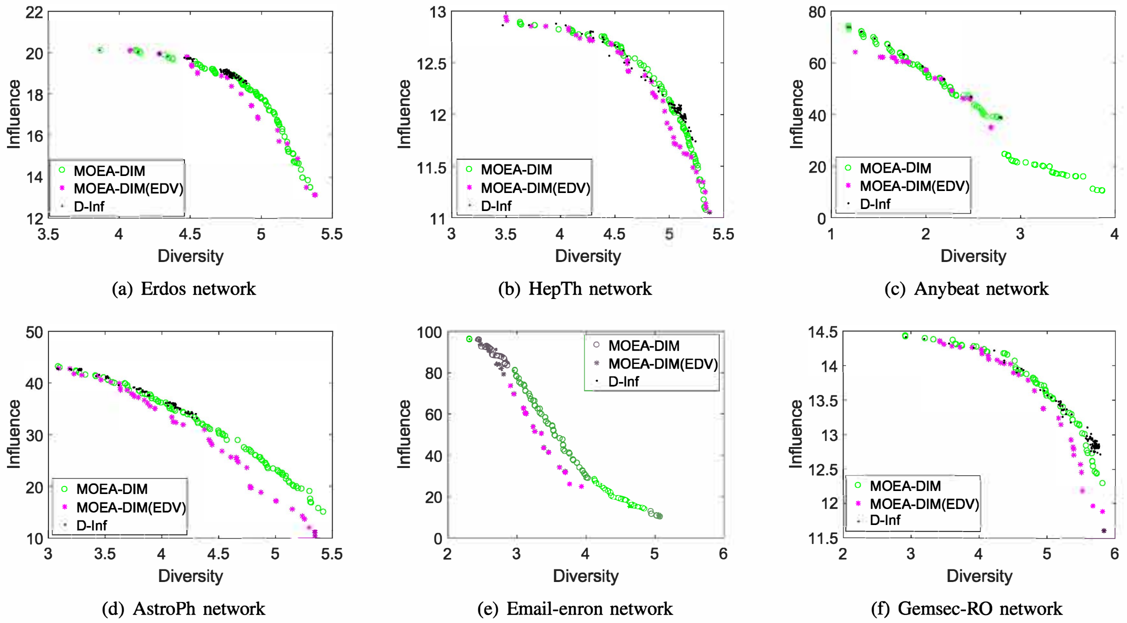
In our experiments, Monte Carlo simulation for 10,000 times is adopted to approximate the influenced crowd under the independent cascade (IC) model [4] for all datasets. All the experiments are conducted in Java on an Intel(R) Core i7 computer with 3.40 GHz and 20.00 GB Memory. The experimental parameters of comparison algorithms are listed in Table II.

B. Effectiveness of The Proposed Algorithm MOEA-DIM

Fig. 2 shows the plots of final obtained solutions in objective space for comparison algorithms on the six networks under the IC model (the size of seed is set to 10), where the x-axis represents the diversity distribution and the y-axis represents the influence distribution estimated by running 10,000 times of Monte Carlo simulations. Note that D-Inf only obtains one solution when the weighting parameter γ is fixed. In order to get multiple solutions for D-Inf, γ ranging from 0 to 1 is equally divided into 100 values, since that there are about 100 non-dominated solutions obtained by our algorithm for most of datasets.

From this figure, for MOEA-DIM, there exists a tradeoff between objectives *influence* and *diversity* on the six networks. In other words, the two objectives influence and diversity conflict with each other in some degree, this is the reason why a multi-objective evolutionary algorithm is proposed to optimize influence and diversity simultaneously. In addition, we can find that the final solutions obtained by D-Inf are usually concentrated on the upper half of the Pareto front of MOEA-DIM on most of networks. Compared with MOEA-DIM, the D-Inf algorithm is poorly distributed. MOEA-DIM(EDV) is similar to MOEA-DIM in the distribution trend. However, the non-dominated solutions obtained by MOEA-DIM(EDV) are most dominated by that obtained by MOEA-DIM on the six datasets, due to the fact that the influence is approximated by the method of EDV in MOEA-DIM(EDV).

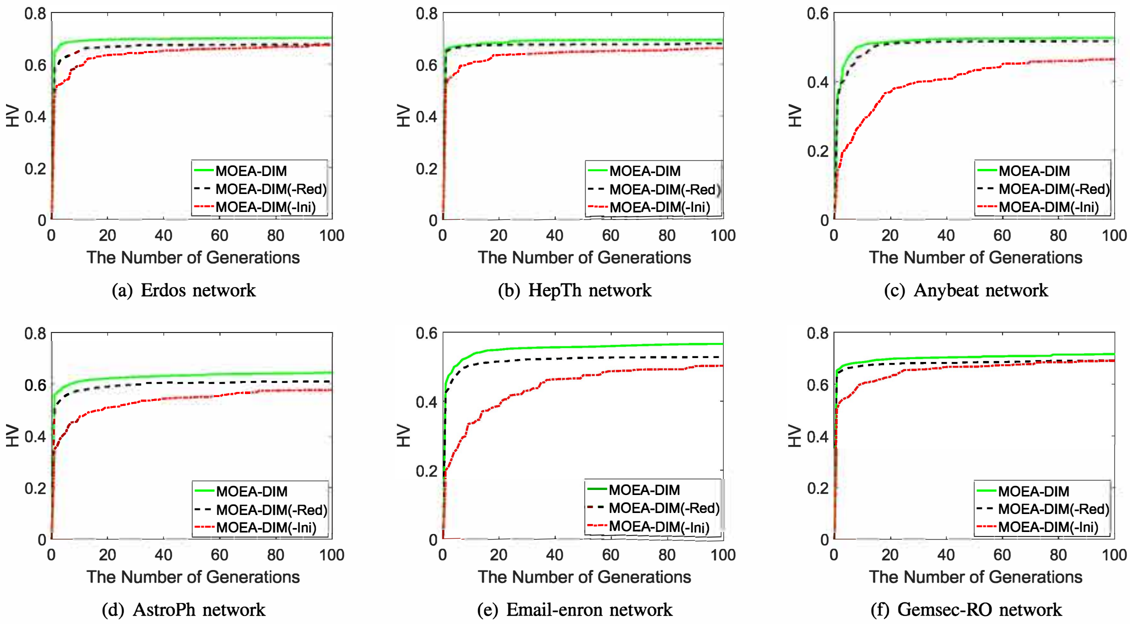
In addition, Table III shows the running time of D-Inf, MOEA-DIM and MOEA-DIM(EDV) on the six networks. From this table, it can be found that the proposed algorithm MOEA-DIM is much faster than the greedy algorithm D-Inf



(a) Erdos network (b) HepTh network (c) Anybeat network

(d) AstroPh network (e) Email-enron network (f) Gemsec-RO network

Fig. 2. The plots of final obtained solutions in objective space for comparison algorithms on the six networks (when the size of seed is set to 10).



(a) Erdos network (b) HepTh network (c) Anybeat network

(d) AstroPh network (e) Email-enron network (f) Gemsec-RO network

Fig. 3. The hypervolume (HV) of non-dominated solutions obtained by MOEA-DIM, MOEA-DIM(-Red) and MOEA-DIM(-Ini) with different generations on the six networks (when the size of seed is set to 10).

on all networks, but is much slower than MOEA-DIM(EDV) since that MOEA-DIM(EDV) use an approximate evaluation method to calculate the influence by replacement of the Monte Carlo simulation. Therefore, the effectiveness and efficiency of the proposed algorithm MOEA-DIM compared with baselines are verified.

C. *Effectiveness of The Proposed Strategies Used in MOEA-DIM*

In this section, we will verify the effectiveness of the proposed strategies used in MOEA-DIM, that is, the network reduction strategy and population initialization strategy. Fig. 3 shows the hypervolume (HV) [31] (The larger values of hypervolume, the better the solutions found in objective space) of non-dominated solutions obtained by MOEA-DIM, MOEA-DIM(-Red) and MOEA-DIM(-Ini) with different generations on the six networks (the size of seed set is set to 10), where x-axis represents the number of population iterations and the y-axis represents the the hypervolume values. As can be found from this figure, the proposed algorithm MOEA-DIM can converge very quickly within 20 generations on most of networks. It can be also observed that MOEA-DIM with the proposed two strategies can converge faster and better than MOEA-DIM(-Ini) and MOEA-DIM(-Red). Therefore, the effectiveness of the proposed two strategies used MOEA-DIM is verified.

V. Conclusion and Future Work

In this paper, we formulated the task of diversified social influence maximization as a multi-objective problem and then proposed a multi-objective algorithm named MOEA-DIM for it. To be specific, we employed the overlapping community detection method to obtain the community structure information of nodes, by which a novel diversity measure was designed without using any node's category information. In MOEA-DIM, a network reduction strategy based on overlapping community structure information was designed to greatly reduce the search space. In addition, a population initialization strategy based on random walk was suggested to accelerate the convergence of the algorithm. Finally, experimental results on six real-world networks with different characteristics demonstrated that the proposed MOEA-DIM algorithm has good performance in terms of both effectiveness and efficiency. The proposed MOEA-DIM has shown promising performance in diversified social influence maximization on medium-sized networks. In the future, we would like to design more efficient MOEAs for diversified social influence maximization on very large-scale networks.

Acknowledgement

This work is supported by the Key Project of Science and Technology Innovation 2030 supported by the Ministry of Science and Technology of China (Grant No. 2018AAA0101300), the Natural Science Foundation of China (Grant No.61976001 and 61876184), and Humanities and Social Sciences Project of Chinese Ministry of Education (Grant No.18YJC870004).

References

[1] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, and J. H. Fowler, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, 2012.

[2] E. Cambria, H. Wang, and B. White, "Guest editorial: Big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 1–2, 2014.

[3] M. Grassi, E. Cambria, A. Hussain, and F. Piazza, "Sentic web: A new paradigm for managing social media affective information," *Cognitive Computation*, vol. 3, no. 3, pp. 480–489, 2011.

[4] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2003, pp. 137–146.

[5] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 420–429.

[6] A. Goyal, W. Lu, and L. V. Lakshmanan, "Celf++: optimizing the greedy algorithm for influence maximization in social networks," in *Proceedings of the 20th international conference companion on World wide web*. ACM, 2011, pp. 47–48.

[7] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

[8] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in *European conference on principles of data mining and knowledge discovery*. Springer, 2006, pp. 259–271.

[9] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 1029–1038.

[10] M. Gong, C. Song, C. Duan, L. Ma, and B. Shen, "An efficient memetic algorithm for influence maximization in social networks," *IEEE Computational Intelligence Magazine*, vol. 11, no. 3, pp. 22–33, 2016.

[11] Q. He, X. Wang, Z. Lei, M. Huang, Y. Cai, and L. Ma, "Tifim: A two-stage iterative framework for influence maximization in social networks," *Applied Mathematics and Computation*, vol. 354, pp. 338–352, 2019.

[12] A. Talukder, M. G. R. Alam, N. H. Tran, D. Niyato, and C. S. Hong, "Knapsack-based reverse influence maximization for target marketing in social networks," *IEEE Access*, vol. 7, pp. 44 182–44198, 2019.

[13] L. Jiang, X. Zhao, B. Ge, W. Xiao, and Y. Ruan, "An efficient algorithm for mining a set of influential spreaders in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 516, pp. 58–65, 2019.

[14] H. Huang, H. Shen, Z. Meng, H. Chang, and H. He, "Community-based influence maximization for viral marketing," *Applied Intelligence*, vol. 49, no. 6, pp. 2137–2150, 2019.

[15] F. Tang, Q. Liu, H. Zhu, E. Chen, and F. Zhu, "Diversified social influence maximization," in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE, 2014, pp. 455–459.

[16] X. Zhang, F. Duan, L. Zhang, F. Chen, Y. Jin, and K. Tang, "Pattern recommendation in task-oriented applications: a multi-objective perspective," *IEEE Computational Intelligence Magazine*, vol. 12, no. 3, pp. 43–53, 2017.

[17] L. Zhang, G. Fu, F. Cheng, J. Qiu, and Y. Su, "A multi-objective evolutionary approach for mining frequent and high utility itemsets," *Applied Soft Computing*, vol. 62, pp. 974–986, 2018.

[18] L. Zhang, H. Pan, Y. Su, X. Zhang, and Y. Niu, "A mixed representation-based multiobjective evolutionary algorithm for overlapping community detection," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2703–2716, 2017.

[19] X. Zhang, K. Zhou, H. Pan, L. Zhang, X. Zeng, and Y. Jin, "A network reduction-based multiobjective evolutionary algorithm for community detection in large-scale complex networks," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 703–716, 2020.

[20] L. Zhang, J. Xia, F. Cheng, J. Qiu, and X. Zhang, "Multi-objective optimization of critical node detection based on cascade model in complex networks," *IEEE Transactions on Network Science and Engineering*, vol. In press, 2020.

[21] J.-R. Lee and C.-W. Chung, "A fast approximation for influence maximization in large social networks," in *Proceedings of the 23rd international conference on World Wide Web*. ACM, 2014, pp. 1157–1162.

[22] A. Caliò, R. Interdonato, C. Pulice, and A. Tagarelli, "Topology-driven diversity for targeted influence maximization with application to user engagement in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2421–2434, 2018.

[23] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[24] J. Xie, B. K. Szymanski, and X. Liu, "Slpa: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *2011 IEEE 11th international conference on data mining workshops*. IEEE, 2011, pp. 344–349.

[25] T. Lou and J. Tang, "Mining structural hole spanners through information diffusion in social networks," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 825–836.

[26] S. Kim, D. Kim, J. Oh, J.-H. Hwang, W.-S. Han, W. Chen, and H. Yu, "Scalable and parallelizable influence maximization with random walk ranking and rank merge pruning," *Information Sciences*, vol. 415, pp. 171–189, 2017.

[27] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI*, 2015. [Online]. Available: http://networkrepository.com

[28] [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/data

[29] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, pp. 2–es, 2007.

[30] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie, "Simulated annealing based influence maximization in social networks," in *Twenty-fifth AAAI conference on artificial intelligence*, 2011, pp. 7–11.

[31] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithmsła comparative case study," in *International conference on parallel problem solving from nature*. Springer, 1998, pp. 292–301.