

Solving the Influence Maximization-Cost Minimization Problem in Social Networks by Using a Multi-Objective Differential Evolution Algorithm

Peng-Li Lu², Li Zhang², Jian-Xin Tang^{1,2*},
Ji-Mao Lan², Hong-Yu Zhu², Shi-Hui Song²

¹ Wenzhou Engineering Institute of Pump&Valve, Lanzhou University of Technology,
Wenzhou, 325100, China
tangjx@lut.edu.cn

² School of Computer and Communication, Lanzhou University of Technology,
Lanzhou 730050, China
{lupl, zhangli20, lanjm, zhuhy, songsh}@lut.edu.cn

Received 11 December 2022; Revised 20 March 2023; Accepted 1 May 2023

Abstract. Influence maximization (IM) problem in social network analysis aims to select a set of the most influential users that can maximize the influence spread in a network. The existing majority of efforts merely focus on the purpose of maximizing the spread of influence. Whereas the budget cost is a major factor needed to be taken into consideration in practical scenarios. In this paper, we consider both the influence maximization and the cost minimization simultaneously in the process of influence spreading, and formulate the two targets as a multi-objective combinational optimization problem. A discrete multi-objective differential evolution optimization (DMODE) with mutation, crossover and selection operators specifically for the topological network structure is proposed. The algorithm combines multiple mutation operators to enhance exploration and exploitation, and an exploiting strategy based on degree ranking is developed to improve the convergence performance. Numerous experiments on four real-world social networks are conducted, and the obtained results demonstrate the outperformance of the proposed algorithm over the state-of-the-art methods.

Keywords: social network, multi-objective optimization, influence maximization, cost minimization, discrete differential evolution algorithm

1 Introduction

Social networks are playing significant roles in the communication of our daily life as well as many other social activities such as marketing campaigns of innovative products or services, political movements, etc. Billions of individuals have plunged into Twitter, TikTok, Facebook, WeChat and Instagram, which have become primary social network platforms for the fast dissemination of the news, novel products, opinions and knowledge among the users. The intricate interactivity among the individuals in the social networks has aroused the research on viral marketing [1, 2], social behavioral analysis [3], rumor control [4], revenue maximization [5], opinion formation [6] and collective decision-making [7, 8].

Identifying influential spreaders from the network is one of the most significant topics of social network analysis. And seeking an influential node set of given size with the most collective spread of influence is regarded as influence maximization problem [15], which was modeled as a discrete optimization problem by Kempe et al. [9]. It has aroused the great interest of researchers due to the wide applications of influence maximization problem [10]. Viral marketing is one of such practices, of which the pivotal aim is to maximize the dissemination of promotion on novel products by offering free or discount samples to a set of initial consumers with the hope of recommending it to their adjacent friends in a network perspective [11].

Budget cost serves as an essential factor in various marketing applications. However, the majority of existing IM solutions merely take the influence maximization into account while ignoring the cost minimization. For instance, a salesman wants to market and popularize a product, he/she would like to provide the product available for free use to some consumers with high reputations. These adopters will intensely spread information about the product to their friends and the friends of their friends, through which the marginal gain can be maximized. So

* Corresponding Author

the salesman has to consider the total cost of employing these influential consumers under a limited budget, and such a scenario can be expressed as a multi-objective influence maximization problem. The corresponding target of the problem is to select a set of seeds that can maximize the amount of active nodes and minimize the budget cost simultaneously.

A few recent studies have considered the multi-objective variants of the IM problem. Mohammadi and Saracee explored the relationship between influence maximization and the time factor, striving to achieve the maximum range of influence propagation in the least amount of spreading time [44]. Since the proximity of a node in the network to the graph center is a valid measure of propagation capacity, Sheikhamadi and Zareie [12] incorporated the K -shell value in solving the influence maximization problem. Further, the budget is used as a constraint to maximize both the influence propagation function and the K -shell value of the seed set. Maximizing sales volume while reducing costs is a multi-objective problem in actual marketing campaigns [13]. However, the limited existing work attempted to address the bi-objective IM problem that seeks to maximize the effect of seed sets while minimizing the selection costs [14, 36]. Those presented algorithms keep the number of seed nodes unchanged when solving multi-objective influence maximization problem, but in actual marketing, the seed set size may vary from the budget. So it is not practical to arbitrarily treat the number of seed nodes as a constant parameter and then identify the fixed number of influential nodes at a lower cost. The above problem has been noticed by literatures [37, 53], in which the authors set the number of seed nodes to be a variable and proposed a bi-objective IM problem that maximizes the influence propagation while considering minimizing the budget cost. Nevertheless, the algorithms are time-consuming on large-scale networks, posing that there are still great challenges in identifying the multifarious solutions to the multi-objective influence maximization problem. Therefore, investigating the relationship between the seed set size and influence spreading coverage, and developing more effective policies for the problem needs to be further explored.

The differential evolution algorithm (DE) [50] has been proved to be one of the most robust evolutionary algorithms, and has been widely used in various fields because of its simplicity, effectiveness and few control parameters [35]. In order to find diverse Pareto non-dominated solutions, in this paper, an improved differential evolution algorithm with multiple mutations, crossover and selection operators specifically for the topological network structure is put forward to address the influence maximization-cost minimization (IM-CM). Numerous experimental simulations on social networks validate that the presented algorithm is effective in terms of the diversity of the Pareto non-dominated solutions. The main contributions of this work can be emphasized as follows:

- (1) A discrete differential evolution optimization based on network topology structure is proposed for the IM-CM problem.
- (2) The proposed DMODE algorithm conceives multiple mutation operators to strengthen the exploration and exploitation in the discrete topological solution space.
- (3) An exploitation strategy based on degree ranking is designed for the mutation operation to improve the local search performance of the algorithm.
- (4) Empirical results on four real-world networks demonstrate that the proposed discrete differential evolution algorithm is effective and efficient in optimizing the multi-objective influence maximization, and can be scalable to large-scale networks.
- (5) The application scenarios of the intelligence evolutionary algorithm are further enriched by mapping it into topological network space.

The rest of this paper is structured as follows. Previous works on the influence maximization problem is presented in Section 2. Section 3 introduces the concepts of IM and the original differential evolution algorithm. The proposed algorithm is detailed in Section 4. Section 5 describes the main results of the simulations and the performance comparison with other algorithms. Section 6 summarizes the work with future work.

2 Related Works

2.1 Greedy-Based Methods

Domingos and Richardson [15] initially explored the marginal gain in viral marketing from the viewpoint of network and treated the problem as an influence maximization. Kempe et al. formulated the problem as a discrete optimization problem and proved it to be NP-hard [9]. The authors employed the Monte-Carlo mechanism to approximate the marginal return of each node in the network and selected the best node in each round using a

greedy algorithm based on the hill-climbing strategy. Although greedy algorithm can always give an approximation of 63% of the optimal solution, it is time-consuming due to the need for thousands of Monte-Carlo simulations for each estimation, especially in large scale networks.

To relieve the high time consumption of the original greedy strategy, Leskovec et al. [17] explored the properties of the submodular function and presented the Cost-Effective Lazy Forward (CELF) algorithm. A further optimized version called CELF++ was presented by Goyal et al. [18], which showed 50% efficiency improvement over the CELF in some networks. Yu et al. [46] designed a three-stage greedy algorithm based on user attributes. The algorithm firstly deletes useless nodes in the network, then puts the remaining nodes into the group of candidate seeds, and finally selects the seed set with the largest influence coverage from the candidate set. Experimental results proved that the accuracy of the solution found by the method is high. Following the greedy-based methods, subsequent improved algorithms were reported in literature [19, 20]. Although there is greatly improvement in reducing time consumption when dealing with large-scale networks, the greedy-based algorithms still suffer from low efficiency and even the accuracy of solutions.

2.2 Heuristic Methods

Topological centrality metrics were widely adopted to identify influential nodes for the IM problem. Besides the high degree centrality, betweenness centrality and other intuitive topology-based methods, Wang et al. [21] proposed a novel centrality algorithm to mine key nodes by combining K -shell and information entropy. Fei et al. [22] proposed an inverse square law based on propagation paths to determine influential nodes. Abdollahpouri et al. [23] formulated a node centrality method by using the local topology properties as well as the global information of the nodes, in which only the sum of the shortest paths of some special nodes was calculated, making it less complicated than degree centrality and closeness centrality. Qian et al. [33] suggested an innovative measure named clone selection based influence maximization algorithm inspired by Clonal Selection Theory. Eigenvectors centrality was adopted to rank the individuals in the population to enhance the performance of the algorithm. A heuristic method considering redundancy weakening and two types of seeds including the loose neighbors and the close neighbors into degree discount was introduced by Wang et al. [24]. When employed to large-scale networks, the heuristic method can reduce the computation time.

A learning algorithm was introduced by Manchanda et al. [47] to handle combinatorial optimization problems through using a node filtering method. There are emerging researches [48, 49] adopting graph neural networks for influence maximization recently. Once the network topology is given, the set of seed nodes based on the topological measures is determined directly. However, extensive simulations showed that the spread of influence of the set of seed nodes identified by the centrality measures tends to be overlapped, and is always inferior to that of greedy-based methods.

2.3 Meta-Heuristic Methods

In the last few years, the superior performance of meta-heuristic algorithms on identifying influential nodes were demonstrated by optimizing a set of combinatorial objective functions. Simulated annealing was the first to be applied by Jiang et al. [25] to solve influence maximization, where the Monte-Carlo simulation mechanism was substituted by a function named expected diffusion value to evaluate the potential local influence of given candidate nodes. The meta-heuristic can greatly decrease the complexity of influence evaluation of given nodes, but tends to fall into local optimum easily. Gong et al. [16] proposed to approximate the expected influence propagation within two-hop neighbors of the candidate node and introduced a discrete particle swarm optimization (DPSO) by modeling the evolutionary rules of the particles based on discrete network topology. Experiments showed that DPSO is superior to the greedy-based CELF++ in the perspective of seed quality. Cui et al. [26] introduced a degree descending search evolution (DDSE), in which the operations of the differential evolution including mutation, selection and crossover were adopted to update the candidate sets. However, DDSE performs ineffective in many scenarios due to its same parameter settings for each network. Considering the transmission of information or products does not have only one single property, Ni et al. [27] investigated the spread pattern of information in the way of simultaneous as well as independent in social networks from the perspective of multi-attribute of influence transmission, so as to obtain the approximate guarantee of the optimal solution.

Recently, researchers successfully applied the ant colony optimization [31], gray wolf optimization [32], moth-flame optimization [28] as well as the differential evolution approach [29] to deal with influence maxi-

zation problem and demonstrated the promising of meta-heuristics. In addition, the dynamic solutions were designed to better fit the IM problem, due to the evolution of the social network topology over the time [30].

2.4 Multi-Objective Influence Maximization

Generally, the problem of maximizing influence is affected by various factors. To depict the actual situations, Yang et al. [14] took into account the cost minimization while solving the influence maximization and formulated it as a multi-objective IM-CM problem.

A multi-objective biogeography-based strategy that optimizes both the influence maximization and cost minimization functions was presented by Sagar et al. [38]. The algorithm combines the optimal properties of biogeography-based strategy with non-dominated sorting genetic algorithm-II (NSGA-II) to reduce the convergence time. Sheikahmadi and Zareie [39] defined the multi-objective function as an influence metric, and used artificial bee colony approach to find near-optimal solutions within an acceptable time duration. A novel framework named PRNSGA-II based on NSGA-II was suggested by Qian et al. [36] to address IM-CM. The algorithm used PageRank value as an objective function, which can be utilized to sort and identify important nodes to effectively expand the coverage of influence dissemination. Olivares [37] adopted three swarm intelligence algorithms i.e., particle swarm optimization (PSO), black hole optimization (BHO) and bat algorithm (BA) to handle the IM-CM under the linear threshold model, and the experimental outcomes illustrated the three algorithms show greater convergence and higher diversity of solutions in resolving the multi-objective IM problem.

To reduce the running time, Jia et al. [40] defined IM-CM problem as minimum cost seed selection with probabilistic influence spreading guarantee and an algorithm called Sampling-Greedy was put forward. A constraint bound that changes over time was considered by Roostapour et al. [41] into the selection of the most influential node. The authors utilized the Pareto optimization method to achieve linear time approximation. Cunegatti et al. [42] devised a multi-objective evolutionary algorithm (MOEA) based on network scaling. The approach first downsizes the network to find the most influential nodes, and then maps these nodes to the original network by heuristic methods. The algorithm is effective in processing multi-objective problems, and the running time is improved.

The multi-objective influence maximization has two or more objectives conflicted with each other needed to be optimized. Searching all the non-dominated solutions is therefore essential for a multi-objective algorithm. The challenge of multi-objective influence maximization problem is how to well balance the influence spread maximization between other targets, such as budget cost [43] or time consumption [44], and obtain a set of accuracy Pareto solutions. Therefore, developing effective and efficient algorithms for the problem remains a challenging research topic. In this paper, the differential evolution optimization, which has been widely applied in complex continuous optimization problems, is employed to address the IM-CM problem effectively.

3 Preliminaries

3.1 Influence Maximization and Propagation Model

Given a social network $G=(V, E)$, identifying and selecting k individuals, which have the greatest ability to spread the information, as primary active nodes from nodes set V into the seed set S is the main purpose of the IM problem.

$$S^* = \arg \max_{S \subseteq V, |S|=k} \sigma(S), \quad (1)$$

where S^* is the optimal seed set.

In this paper, we employ the LT model to simulate the influence spreading in social networks. Linear threshold model is an influence cumulative model. Under the model, every node has a particular critical value θ that represents the difficulty of the node will be affected by its adjacent active neighbors. The lower the threshold is, the higher the potential that the node can be affected. The nodes that are in the active state at step t have chances to activate its inactive neighbors. During the process, the influence received by each inactive node is continuously accumulated. An inactive node v updates its state into active if and only if its threshold is exceeded by the sum of

the received influence, and it remains active till the spreading process ends.

As defined in [37], given a directed and weighted graph $G = (V, E, W, h)$ with node set $V = \{v_1, v_2, \dots, v_n\}$, edge set $E = \{e_1, e_2, \dots, e_m\}$ and edge weight vector $W = \{w_1, w_2, \dots, w_m\}$, h represents a threshold function which allots one threshold θ to every node. Given a group of initial nodes $S \subseteq V$, the set of active nodes is $F(S) \subseteq V$. Equation (2) can be used to express the model.

$$F_t(S) = F_{t-1}(S) \cup \left\{ v \in V \mid \sum_{\{u \in F_{t-1}(S) \mid (u, v) \in E\}} w(u, v) \geq h(v) \right\}, \quad (2)$$

where $F_t(S)$ denotes the active node set at step t , and $w(u, v)$ is the influence of node u transmitted on v , i.e., the weight on the edge (u, v) . At the very beginning $t = 0$, merely the seed nodes are active, thus $F_0(S) = S$. Then, at time $t > 0$, all the nodes in $F_{t-1}(S)$ plus the nodes that are activated at time t constitute the active node set. In the linear threshold model, each node whose threshold is less or equal than the total weight of the edges from the nodes in $F_{t-1}(S)$ pointing to it will be added to the active nodes set. For convenience, equation (2) is used as the first objective (i.e., influence maximization function) of IM-CM problem in this paper.

3.2 Multi-objective Optimization

A problem that optimizes multiple objective functions simultaneously is named multi-objective optimization problem. Hence the objective functions can be represented through a multidimensional optimization vector, as formulated in equation (3).

$$\begin{aligned} \maximize \quad & F(x) = [f_1(x), f_2(x), \dots, f_k(x)]^T, \\ \text{s.t.} \quad & x = (x_1, x_2, \dots, x_n) \in \Omega, \end{aligned} \quad (3)$$

where $f(x)$ is one of the k objective functions to be optimized, $x = (x_1, x_2, \dots, x_n)$ is a potential solution with n variables and Ω is the multi-dimensional decision space of the problem.

Pareto dominance and Pareto solutions are the fundamental components in solving multi-objective problems. Given two decision making variables $x_A, x_B \in \Omega$, if and only if $f_i(x_A) > f_i(x_B)$ and $f_i(x_A) \neq f_i(x_B)$, $i = 1, 2, \dots, k$, then the x_A dominates x_B , expressed as $x_A \prec x_B$. Let the decision variable vector $x \in \Omega$ be a Pareto non-dominated solution, if there not exists a decision variable $x^* \in \Omega$ which dominates decision variable x , then the group of all non-dominated solutions is viewed as the Pareto solution set. The objective function values of every Pareto solution constitute the Pareto frontier. The mathematical expression is shown in equation (4).

$$PS = \left\{ x \in \Omega \mid \neg \exists x^* \in \Omega, x^* \prec x \right\}. \quad (4)$$

The IM-CM problem mainly consists of two objectives. The first one is influence maximization and the second one is to minimize the cost function. On behalf of defining the cost function, we assume that the cost of every individual in the social networks is the same and the amount of seed nodes is expressed as the cost value. Correspondingly, the formula of cost function is represented in equation (5).

$$\begin{aligned} \text{Cost}(S) &= \sum_{i=1}^n x_i, \\ x_i &\in \{0, 1\}, \forall i \in \{1, \dots, n\}. \end{aligned} \quad (5)$$

How to select the most influential individuals with the least cost is a major work of IM-CM. Based on equation (2) and equation (5), the problem can be formulated as in equation (6).

$$\begin{aligned} f(S) &= \{ \max F(S), \min \text{Cost}(S) \}, \\ \text{s.t.} \quad & |S| \leq |F(S)|. \end{aligned} \quad (6)$$

Obviously, single-objective optimization methods can hardly tackle this problem effectively, as $Cost(S)$ always increases with the increase of $F(S)$. Therefore, $Cost(S)$ is added with a negative sign, and then the goals of the multi-objective function Pareto maximize the influence diffusion function and the negative cost function simultaneously, as reformulated as in equation (7).

$$\max f(S) = \{F(S), -Cost(S)\} . \tag{7}$$

Different from the traditional influence maximization problem with only one objective function that needs to be optimized, IM-CM is a binary optimization problem that takes both the individual’s marginal return and its corresponding cost into consideration. The two-objective problem can produce a set of Pareto solutions, which allow decision makers to make alternative options for different influence spreading situations.

3.3 Differential Evolution Algorithm

Differential evolution is a meta-heuristic global optimization algorithm proposed by Storn and Price [50]. Similar to most of other evolutionary algorithms, the standard DE is composed of mutation, crossover and selection operators. Numerous experiments have proved that DE is one of the most robust algorithm in the evolutionary algorithms variants. Furthermore, as a result of the power of differential evolution algorithm, it has been successfully applied to diversified research fields [45] including project scheduling problem [51], traveling salesman problem [52], and knapsack problem [34].

In the differential evolution optimization, it is supposed that a population consists of NP individuals. Each individual has a D -dimensional space and expressed as in equation (8).

$$X_i = \langle x_i^1, x_i^2, \dots, x_i^D \rangle , \tag{8}$$

where X_i represents the i th individual in the population. The individuals constituting the population evolve according to the operators in a constant number of iterations, eventually obtaining an optimal solution.

4 Proposed Algorithm

4.1 Discrete Encoding Mechanism

There are two kinds of encoding mechanism for the IM problem, including real-number encoding and binary encoding. The former one requires fewer amount of memory consumption but only preserves the form of the key node set. While the latter utilizes a n -bit string to indicate the status of the n nodes in the network, where 1 means the corresponding node has been selected as a seed, and 0 is not.

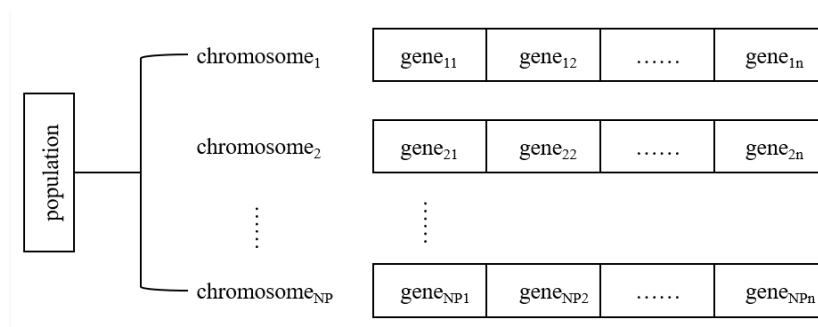


Fig. 1. Population encoding of differential evolution algorithm

According to the IM-CM problem, this paper adopts the binary encoding mechanism to encode the chromosomes. Assumed that each node in the social networks is assigned a distinct positive integer, then a gene in the chromosome can be portrayed by a node in the network. The chromosome population is schematized in Fig. 1, where each gene is marked with 1 or 0, indicating the corresponding node in the node set V is selected as a seed or not. The encoding mechanism is illustrated in Fig. 2.

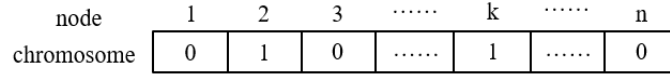


Fig. 2. An encoding illustration of a chromosome

4.2 Framework of DMODE for IM-CM

Initialization, mutation, crossover and selection are four fundamental operations of the discrete multi-objective differential evolution. Fig. 3 gives the flowchart of the algorithm for IM-CM. The framework of DMODE for IM-CM is shown in Algorithm 1.

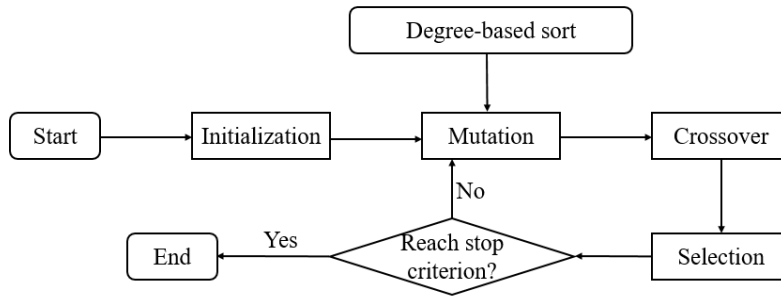


Fig. 3. The flowchart of the DMODE for IM-CM based on degree sorting strategy

Algorithm 1. Framework of DMODE for IM-CM

Require: Graph $G = (V, E, W, h)$, the threshold function h , the crossover probability Cr , the scaling factor F , the number of iteration g_{max} , the population size NP , Candidate node set U and the network size n .

Ensure: Pareto solution set S .

- 1: $X = \text{Initialization}(G, NP)$
 - 2: Select out the best vector X_{best} based on maximum fitness value
 - 3: $t = 1$
 - 4: **while** $t \leq g_{max}$ **do**
 - 5: $M = \text{Mutation}(X, X_{best}, U, F, n)$
 - 6: $C = \text{Crossover}(X, M, n, Cr)$
 - 7: $X = \text{Selection}(X, C)$
 - 8: $S = \max \{F(S), -Cost(S)\}$
 - 9: Update the X_{best}
 - 10: $t = t + 1$
 - 11: **end while**
 - 12: **return** Pareto solution set S .
-

Initialization. Generally, the greater the multiformity of the population is, the higher the solution accuracy becomes. Therefore, the genes in the chromosomes are initialized by non-repetitive nodes selected randomly from the network. The main procedure is outlined in Algorithm 2. An example of initialization population is shown as Fig. 4.

Algorithm 2. Initialization**Require:** Graph $G = (V, E, W, h)$ and the population size NP .**Ensure:** The initial population X .1: Initialize population X with null vectors2: **for** each $X_i \in X (i = \{1, \dots, NP\})$ **do**3: **for** j th dimension ($j = \{1, \dots, |V|\})$ **do**4: $X_{ij} \leftarrow \text{Random}\{0, 1\}$ 5: **end for**6: **end for**7: **return** Initial population X .

	1	2	3	k	n
1	0	1	0	1	0
2	1	0	1	0	1
3	1	1	1	0	0
⋮
NP	0	1	1	0	1

Fig. 4. An illustration of initial population

Mutation. Mutation operation that can determine the quality of the solutions is the core step of DMODE. The mutation process uses the differential operator on two chromosomes to generate a new chromosome. Equation (9) expresses the generation of three independent chromosomes in the mutation process, where X is the original population and NP denotes the population size.

$$\begin{cases} X_{r_1} = X.get(random(NP)) \\ X_{r_2} = X.get(random(NP)) \\ X_{r_3} = X.get(random(NP)) \\ s.t. X_{r_1} \neq X_{r_2} \neq X_{r_3} \end{cases} \quad (9)$$

The syntax structure widely used for mutation section is DE/ $r/s/t$, in which DE indicates differential evolution algorithm, r denotes the mutated vector, s specifies the amount of differential vectors adopted and t represents the crossover mechanism. To improve the performance of the algorithm for exploration and exploitation during optimization, two of the most classic mutation operators (i.e., DE/rand/1/bin and DE/best/1/bin) are employed. The redefined discrete rules of generating mutant individuals are shown in equation (10) and equation (11).

$$M_i = X_{r_1} \oplus F * (X_{r_2} - X_{r_3}) . \quad (10)$$

$$M_i = X_{best} \oplus F * (X_{r_1} - X_{r_2}) . \quad (11)$$

where M_i means the i th mutation chromosome; X_{r_1} , X_{r_2} and X_{r_3} are obtained by equation (9); X_{best} is the chromosome selected based on the maximum fitness value; F is zoom factor, the symbols “ \oplus ” and “ $-$ ” indicate mutation operation and differential operator, respectively.

In equation (10), X_{r_2} and X_{r_3} are employed to conduct differential operation. X_{r_1} is generated as a benchmark chromosome. If the j th gene of X_{r_1} is 1 (i.e., a seed node) and its corresponding node is in the *Candidate* node set generated by the degree-based sorting strategy, as given in Algorithm 4, then it will be replaced by the state of the relative dimensional gene generated by the differential operation on X_{r_2} and X_{r_3} . Equation (11) is similar to equation (10), X_{best} is the basis chromosome, X_{r_1} and X_{r_2} are used to perform differential operation. The mutation process is shown in Algorithm 3. M_{ij} is the state of j th gene of the i th chromosome in the mutant population. The *Sigmoid* function is used to generate the two outcomes, if the function value is greater than a random number, then the gene is represented as 1, and 0 otherwise.

A degree-based sorting strategy that are utilized to accelerate the convergence of the discrete differential evolution algorithm is exploited in the process of mutation. The centrality method returns a set of nodes with relatively high degree. Algorithm 4 describes the specific process of how to obtain this node set in detail.

Algorithm 3. Mutation

Require: The parent population X , the optimal chromosome X_{best} , Candidate node set U , the scaling factor F and network size n .
Ensure: The mutation population X .
 1: Initialize population M with null vectors
 2: Generate X_{r1} , X_{r2} and X_{r3} according to equation (9)
 3: **for** $i = 1$ to $|X|$ **do**
 4: **for** $j = 1$ to n **do**
 5: **if** $\text{rand}(0, 1) \geq 0.5$ **then**
 6: **if** $j \in U \ \&\& \ X_{r1j} == 1$ **then**
 7: $M_{ij} = X_{r1j} \oplus F * (X_{r2j} - X_{r3j})$
 8: **else**
 9: $M_{ij} = X_{r1j}$
 10: **end if**
 11: **else**
 12: **if** $j \in U \ \&\& \ X_{bestj} == 1$ **then**
 13: $M_{ij} = X_{bestj} \oplus F * (X_{r1j} - X_{r2j})$
 14: **else**
 15: $M_{ij} = X_{bestj}$
 16: **end if**
 17: **end if**
 18: **end for**
 19: **end for**
 20: **return** Mutated population M .

Algorithm 4. Degree-based sorting strategy

Require: Graph $G = (V, E, W, h)$.
Ensure: The set of nodes U with relatively high degree.
 1: Calculate the degree of each node in V
 2: Rank all the nodes in V in a degree-descending order, denoted as DA
 3: Count the number of nodes with degree θ as R
 4: **if** $|R| > |V|/3$ **then**
 5: Add the first two-thirds of DA into U
 6: **else**
 7: Add the first two-thirds of DA into U
 8: **end if**
 9: **return** Candidate node set U .

Crossover. The mutant population acquired by the mutation step is prepared for the crossover operation. The operation makes crossover between the mutant chromosomes and the original chromosomes to form the novel chromosome population. Then the chromosome individuals will be utilized in the selection step.

In detail, a stochastic number drawn uniformly from $[0, 1]$ and a preassigned crossover probability Cr are used to determine the crossover population. For every mutated chromosome and its corresponding original chromosome, the proposed method selects the states of genes from mutant into C_{ij} if the Cr is less than a random number, otherwise, the corresponding states from the original chromosome are reserved. The detailed crossover rules are as follows:

$$C_{ij} = \begin{cases} M_{ij}, & \text{if } \text{rand}(0, 1) \leq Cr \\ X_{ij}, & \text{if } \text{rand}(0, 1) > Cr \end{cases} \quad (12)$$

where C_i is the i th crossover chromosome, Cr is the pre-set crossover probability, C_{ij} , X_{ij} and M_{ij} are the states of nodes relative to the j th gene in crossover chromosome C_i , primitive X_i and mutant M_i respectively. Algorithm 5 describes the crossover process.

Algorithm 5. Crossover

Require: The parent population X , the mutant population M , the number of nodes n and the crossover probability Cr .

Ensure: The crossover population C .

```

1: Initialize  $C$  with null vectors
2: for  $i = 1$  to  $|X|$  do
3:   for  $j = 1$  to  $n$  do
4:     Renew  $C_{ij}$  according to equation (12)
5:   end for
6: end for
7: return Crossover population  $C$ .

```

Selection. In the last phase, the selection operation will be performed after finishing crossover. In this step, the proposed algorithm employs fitness value function (i.e., $F(S) - Cost(S)$) to evaluate each chromosome. To be specific, this operation calculates the $F(S)$ and $Cost(S)$ of each crossover chromosome as well as their corresponding original chromosomes. In order to make a tradeoff between the two objective functions, the strategy selects the individual with relatively large $F(S)$ and comparatively small $Cost(S)$ to update the original population. In other words, the original chromosome is substituted by the crossover with optimal fitness value. Through the process, the expected influence of the whole population will be increased iteratively. Specific formula is denoted as equation (13).

$$X_i = \arg \max \{F(X_i) - Cost(X_i), F(C_i) - Cost(C_i)\} . \quad (13)$$

where $F(X_i) - Cost(X_i)$ and $F(C_i) - Cost(C_i)$ mean the maximum gain from the premier individual X_i and the crossover individual C_i , respectively. Finally, the chromosomes with larger fitness values are selected into the population for the next generation. The selection process is explained concretely in Algorithm 6.

Algorithm 6. Selection

Require: The parent population X and the crossover population C .

Ensure: The new population X .

```

1: for  $i = 1$  to  $|X|$  do
2:   Renew  $X_i$  according to equation (13)
3: end for
4: return Next generation population  $X$ .

```

5 Experiments Results

The proposed algorithm is implemented by Python and performed on an Intel(R) Xeon(R) Gold 5218R CPU @ 2.10 GHz with 64 GB memory in a Windows system. Influence spread and Pareto solutions are adopted as two criteria to assess the effectiveness and accuracy of DMODE for influence maximization-cost minimization problem.

5.1 Datasets

To verify the performance of the proposed algorithm, experiments are carried out on four social networks with different statistic characteristics.

- Football. The Football network includes “U.S. high school soccer summer A class regularly season connection figure of 2000”.
- Elegans. Elegans network encompasses the neural network of the *Caenorhabditis elegans* worm (*C.elegans*).
- Wiki-Vote. It is a voting network on Wikipedia.
- p2p-Gnutella04. Nodes denote hosts in the Gnutella peer-to-peer network topology and edges represent links among the Gnutella host computers.

Table 1. Statistical characteristics of the four social networks. $|V|$ is equal to the number of nodes. $|E|$ is the number of edges. $\langle k \rangle$ and d are the average degree and average shortest distance, respectively. C and AC denote the average clustering coefficient and the assortativity coefficient

Networks	$ V $	$ E $	$\langle k \rangle$	d	C	AC
Football	35	118	3.37	2.122	0.339	-0.176
Elegans	306	2345	7.90	2.455	0.292	-0.163
Wiki-Vote	7115	103689	29.15	3.341	0.081	-0.083
p2p-Gnutella04	10876	39994	3.68	4.636	0.006	-0.013

Table 1 introduces the characteristics of the four social networks. Each node of the networks is given a threshold $h(v)$ that is more than half of the influence exerted to it. Given a social network $G = (E, V, W, h)$, for each node $v \in V$, $h(v) = \lfloor w(v)/2 \rfloor + 1$, where $w(v) = \sum \{w(u, v) | (u, v) \in E\}$.

5.2 Baseline Algorithms

The DMODE will be compared with five classic methods including PSO, BHO, BA which are proposed in [37], DC (degree centrality) and NSGA-II.

- PSO: Particle swarm optimization is a meta-heuristic optimization mimicking the behaviors of the flocks.
- BHO: The black hole algorithm simulates the process of black holes continuously absorbing stars and searching for the optimal black hole according to successively randomly generated stars.
- BA: Bat algorithm is a meta-heuristic optimization algorithm that uses the concept of the virtual bats, and employs echolocation to determine their distance from food or obstacles.
- DC: As a local measure, degree centrality is a static way of selecting seed nodes, which firstly sorts the nodes by degree and then selects the top- k seeds with the highest degree.
- NSGA-II: NSGA-II uses the same encoding mechanism as the algorithm in this paper, and adopted the $F(S)$ and $Cost(S)$ functions to maximize the influence propagation, resulting in a set of Pareto non-dominated solutions.

5.3 Simulations

The experimental parameter settings of DMODE are given in Table 2. It's worth noting that all parameter settings related to the baseline algorithms are from the original literature in this paper. The maximal evolutionary generation g_{max} of the five algorithms (DMODE, PSO, BHO, BA and DC) on Football, Elegans, Wiki-Vote and p2p-Gnutella04 are set to 200, 300, 50 and 50. The learning factors $c1$ and $c2$ are set to 1 in PSO, and the inertia weight ω is set to 1. The α and γ are 0.9 in BA, ε is set to 1, and the f_{min} and f_{max} are set to 0.5 and 1.5, respectively.

Table 2. Experimental parameter settings

Parameter	Value	Explanation
F	0.5	Zoom factor
Cr	0.3	Crossover probability
NP	25	Population size

Fig. 5 to Fig. 9 show the convergence plots of the five algorithms on the four real-world networks. In each figure, $|x|$, $|F(x)|-|x|$ and $|F(x)|$ denote the number of seed nodes, the number of nodes activated by the seed set and the total number of nodes that end up in the active status, respectively. It can be seen that the DMODE shows the most satisfying performance on each of the networks from all convergence figures.

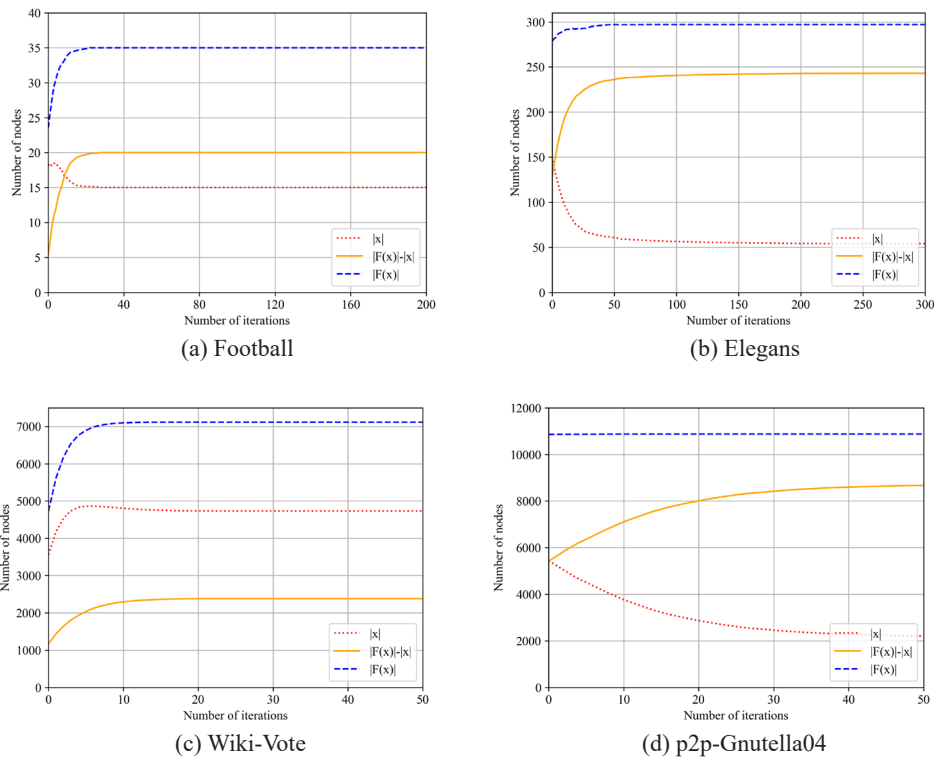


Fig. 5. Evolutionary processes of DMODE on the four networks

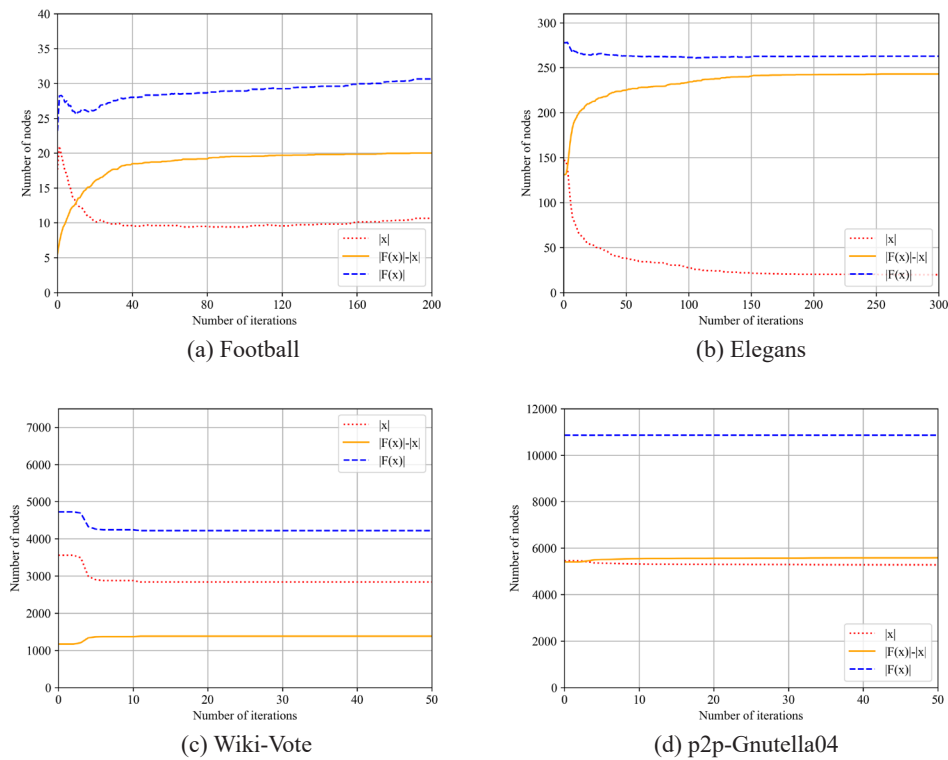
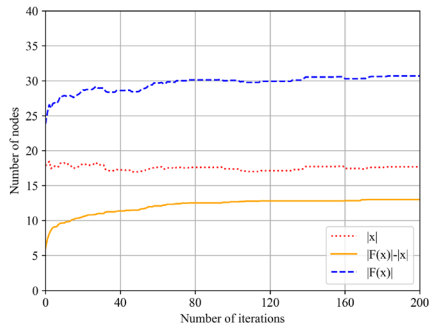
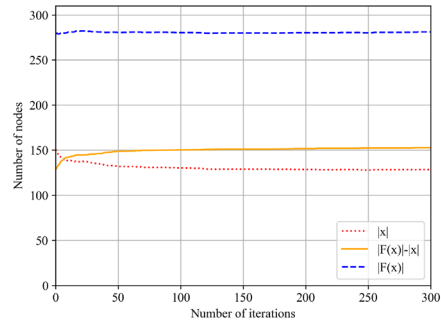


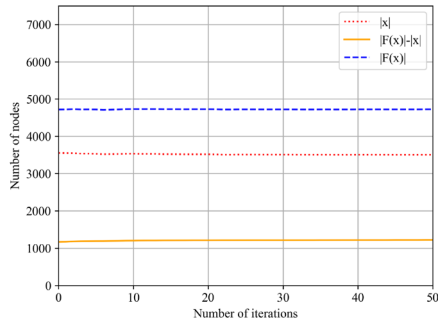
Fig. 6. Evolutionary processes of PSO on the four networks



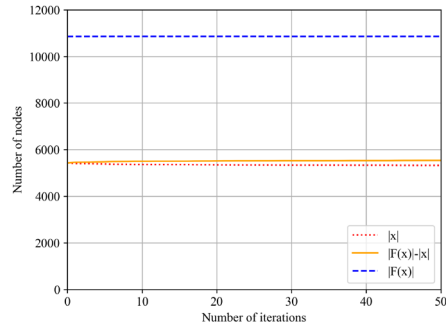
(a) Football



(b) Elegans

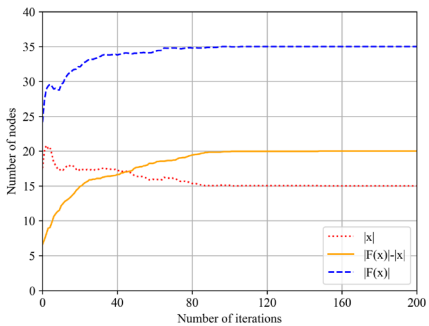


(c) Wiki-Vote

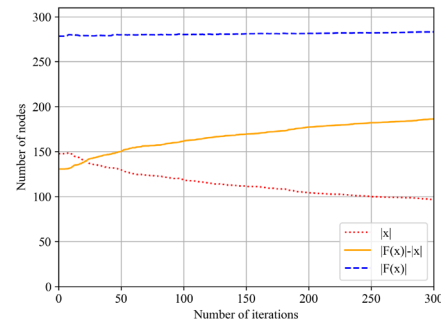


(d) p2p-Gnutella04

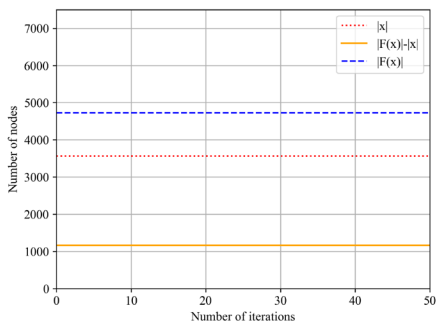
Fig. 7. Evolutionary processes of BHO on the four networks



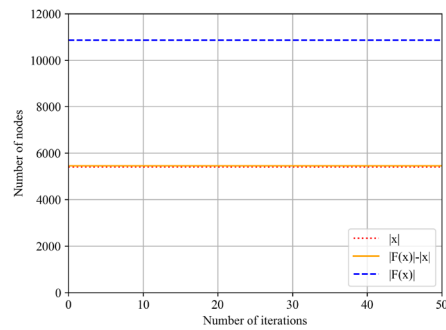
(a) Football



(b) Elegans



(c) Wiki-Vote



(d) p2p-Gnutella04

Fig. 8. Evolutionary processes of BA on the four networks

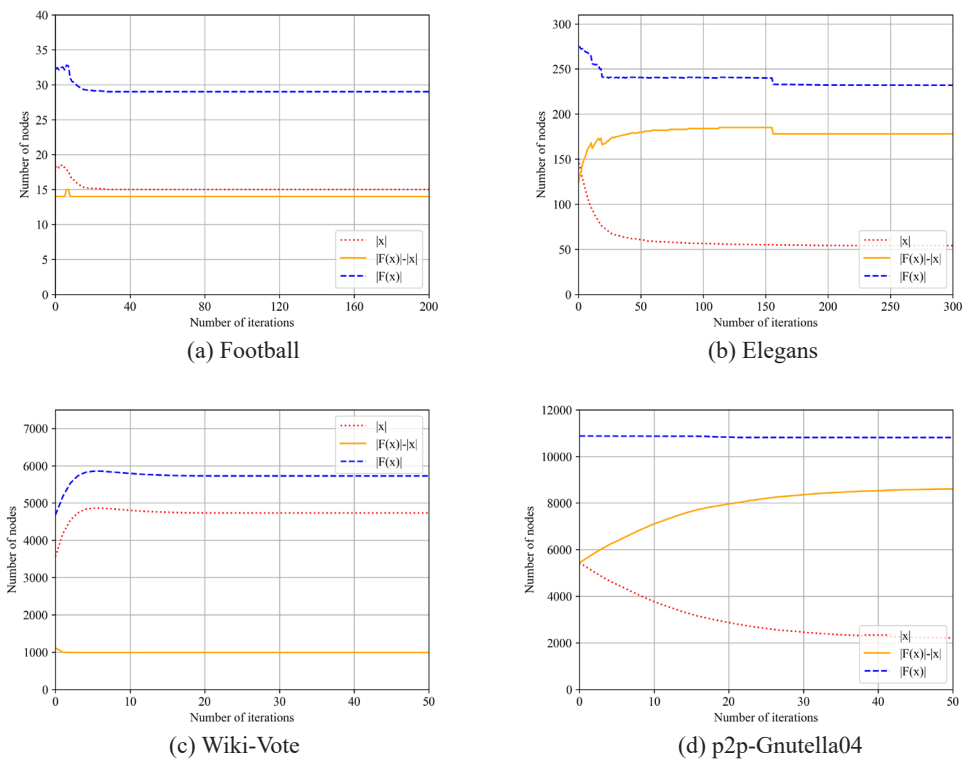


Fig. 9. Evolutionary processes of DC on the four networks

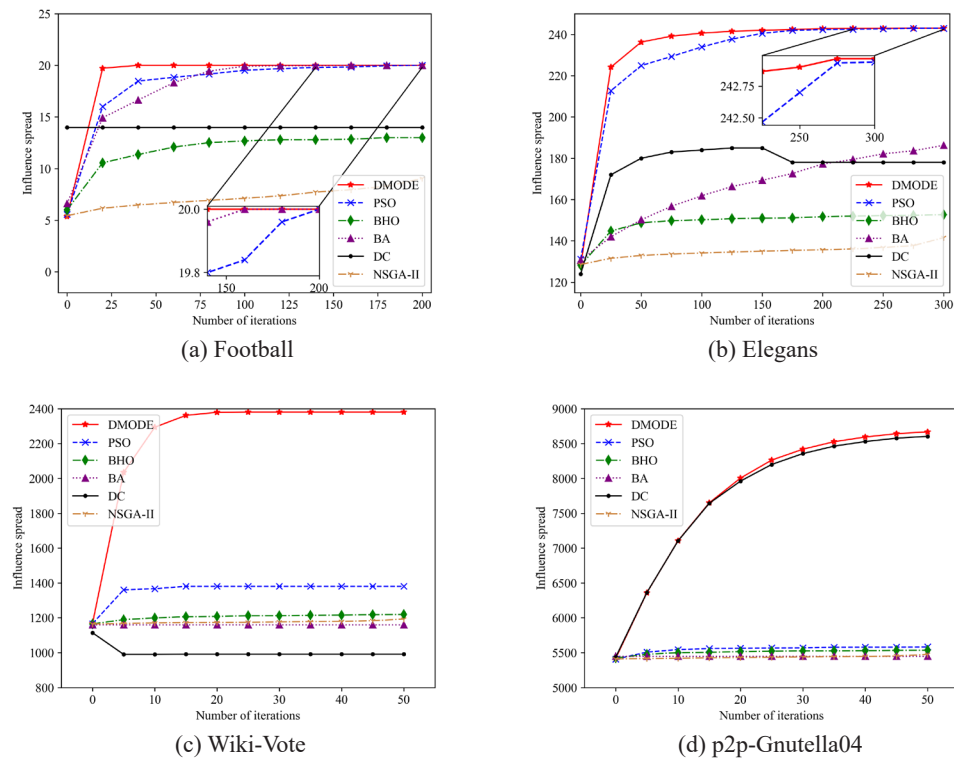


Fig. 10. The influence spread of the algorithms in four datasets

With regard to the Football network, DMODE achieves the maximum coverage of 35 nodes with only 15 seed nodes, while the other four algorithms can hardly reach all the nodes of the network, as shown in Fig. 6(a), Fig. 7(a) and Fig. 9(a). There are a total of 306 users in Elegans, and 245 users are successfully influenced with only 52 seed nodes selected by the DMODE, as shown in Fig. 5(b). Similar results are shown in the Wiki-Vote as well as the p2p-Gnutella04, the superior performance of DMODE is mainly benefiting from the effective evolutionary strategies of DMODE. Fig. 5(c) shows the final result that 2381 individuals in Wiki-Vote are activated by 4734 seed nodes. In the Wiki-Vote network, all five algorithms failed to achieve satisfactory result in terms of influence propagation, but DMODE algorithm ended up with the highest number of active nodes. In short, the DMODE had the largest spreading range compared to the other baseline algorithms.

In Fig. 6, the spreading revenue of PSO on Elegans is more remarkable than the other four algorithms, but it fails to perform well on Wiki-Vote. Compared with DMODE, BHO and BA are mediocre in the four networks, as shown in Fig. 7 and Fig. 8. As can be seen in Fig. 9(d), DC behaves well on p2p-Gnutella04, but yields to PSO, BHO and BA in the other three networks.

5.4 Comparison of Typical Algorithms on IM-CM Problem

The spread performance of DMODE and other five state-of-the-art algorithms (PSO, BHO, BA, DC, NSGA-II) under the LT model is shown in Fig. 10. The curves show that the DMODE is predominant in detecting the targeted seed nodes for IM-CM. The performance of the proposed algorithm is superior to the other baseline algorithms on the four real-world datasets. Due to the degree-based sorting strategy allows chromosome individuals to move quickly towards the optimal position based on the largest fitness value, the influence spread of DMODE grows almost exponentially in the first 20 iterations of each network. Compared to the DMODE, the performance of the other three swarm intelligence algorithms is unstable, they tend to be trapped into premature merely the first few iterations, especially as shown in Fig. 10(c) and Fig. 10(d).

DMODE offers greater advantages in terms of influence transmission coverage compared with NSGA-II. Since NSGA-II is inferior in dealing with multidimensional problems, it becomes incompetent in tackling practical problems in social networks after adopting the mapping mechanism of this paper. With respect to the DC, it is modified to select the top- k nodes with large degree value as seeds, this k is determined by the number of seeds selected by DMODE in the current iteration. As can be seen from Fig. 10, DC turns out to be unstable in tackling the multiple targets at the same time. In Fig. 10(d), since the p2p-Gnutella04 network is large, DC yields results that are competitive with DMODE, however, because the method is highly dependent on the topological properties of the network, leading to overlapping influence, which makes the performance of the algorithm poor. As shown in Fig. 10(b) and Fig. 10(c), DC tends to stagnate after the first few iterations, and even deteriorates.

5.5 Comparison of DMODE with Different Mutation Strategies

In this component, the performance of multiple mutation operators proposed in this paper working jointly is verified. The mutation operation is key part of the proposed algorithm, and the DMODE with two mutation strategies acting together significantly outperforms a single strategy in terms of influence propagation.

DMODE-m is the method that uses solely DE/rand/1/bin. As can be found from Fig. 11, due to the small size of the Football network, the final influence propagation number of this approach is not different from DMODE. In Elegans and p2p-Gnutella04, the number of influence spread generated by DMODE is greater than that of the method utilizing one mutation strategy, the advantage of DMODE is particularly significant in Elegans and Wiki-Vote. As shown in Fig. 11(b), the total influence coverage of DMODE-m does not reach the 297 individuals obtained by DMODE. The algorithm with multiple mutation operators exploited in this paper is superior in the convergence ability of the DMODE and the quality of non-dominated solutions compared to a single mutation method. As can be seen from Fig. 11, although the difference between the two algorithms is not very large, the proposed method still has an advantage in identifying the essential nodes.

5.6 Pareto Front

Pareto front is an important measurement of the algorithms for multi-objective problems. Fig. 12 shows the Pareto fronts of the five meta-heuristics on the four social networks. For the IM-CM problem, each Pareto

non-dominated solution is selected according to the values of $F(S)$ and $Cost(S)$. When Pareto dominance relationship exists between two solution vectors, one of them is chosen as Pareto solution.

As shown in Fig. 12, the Pareto fronts of DMODE are almost the best among the five algorithms. Specifically, in Fig. 12(b) and Fig. 12(c), the Pareto fronts achieved by DMODE are more diverse and stable than the frontiers produced by other algorithms. Due to the small size of the Football network, the Pareto frontiers generated by the five algorithms are relatively concentrated and the gap is not large. As can be seen from Fig. 12(a), Pareto frontiers generated by DMODE algorithm are not dominant, but at a medium level. In terms of the Pareto front, there are some overlaps between the frontiers of DMODE and PSO, but the solutions distribution on the Pareto fronts of DMODE is more uniform, as shown in Fig. 12(b). It can be found from the four networks that the Pareto non-dominated solutions obtained by NSGA-II are more diversity than PSO, BHO and BA, indicating that the method is effective in addressing multi-objective problem. The Pareto front of DMODE on the network p2p-Gnutella04 is slightly fluctuant though it gives a more scattered frontier on the network. In brief, the Pareto solutions of DMODE have a uniform and smooth distribution on the frontiers comparing to the other four algorithms. The stable fronts indicate that the DMODE is a more promising algorithm in resolving the IM-CM problem.

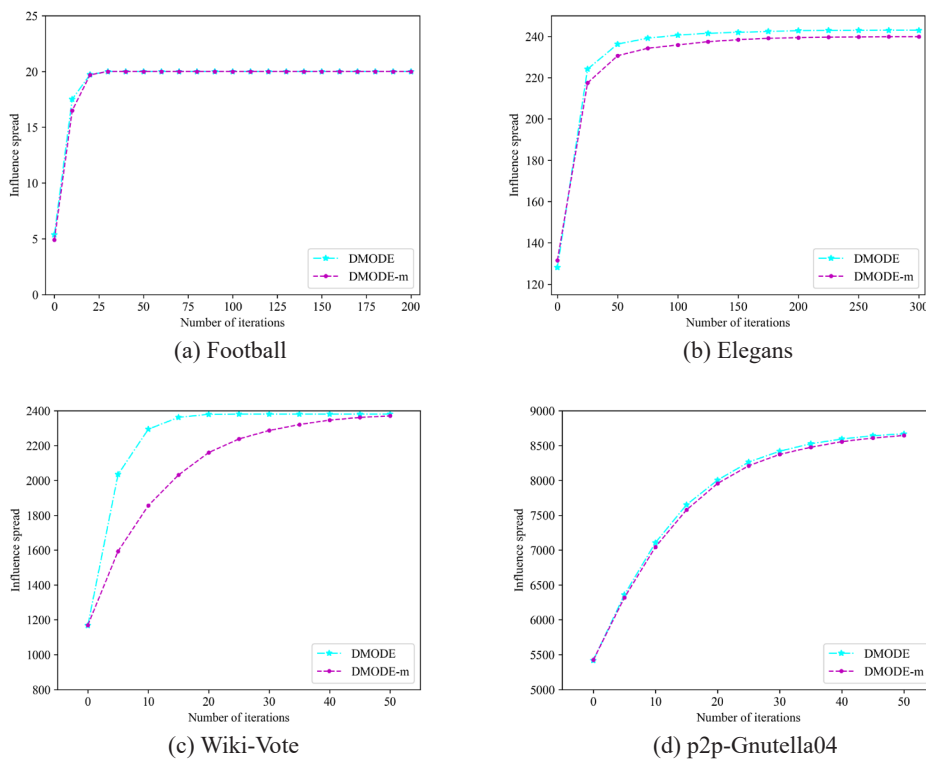
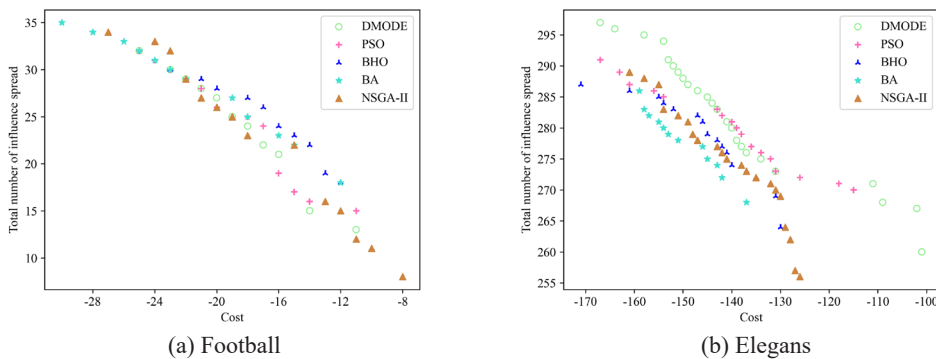


Fig. 11. The influence spread of the algorithms with different mutation operators



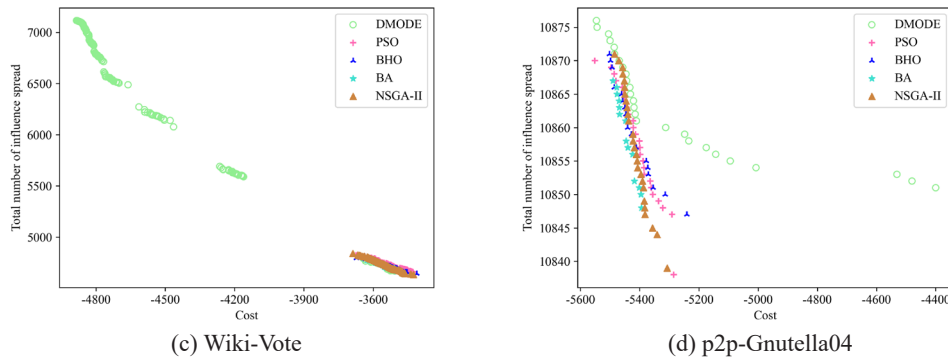


Fig. 12. The Pareto front of the five meta-heuristics on the four social networks

5.7 Statistical Analysis

Since all the algorithms are inherently stochastic, we performed a non-parametric statistical analysis to check whether there is a high level of statistical significance in the results of the algorithms on the IM-CM problem. Firstly, the algorithms were ranked using the Friedman test based on two performance metrics (i.e., the fitness value of the influence maximization objective and the fitness value of the cost minimization objective), as shown in Table 3. The mean rank of each algorithm is derived for each dataset by considering its performance on each setting over multiple runs. Then, the total mean value of each algorithm is calculated, which is the arithmetic mean of the average rank of each algorithm over all the four networks, as shown in the sixth column of Table 3. For simplicity, the overall influence propagation value is utilized as a criterion (in this case, that is the maximization value). The higher the overall mean value, the better the performance of the algorithm. According to the Friedman test results, the proposed algorithm DMODE ranks as the first, followed by NSGA-II, BA, BHO and PSO. In short, the proposed algorithm has a better performance compared to other advanced algorithms.

Table 3. The rank of different algorithms obtained by Friedman test

Algorithm	Football	Elegans	Wiki-Vote	p2p-Gnutella04	Mean value	Rank
DMODE	4.43	4.67	4.32	4.68	4.53	1
PSO	1.14	3.42	1.21	3.25	2.26	5
BHO	2.14	2.33	2.68	2.64	2.44	4
BA	3.57	1.21	4.63	1.36	2.69	3
NSGA-II	3.71	3.38	2.16	3.07	3.08	2

6 Conclusion

In this paper, a novel discrete multi-objective differential evolution optimization with multi-mutation operators, crossover and selection operators specifically for the topological network structure is proposed to tackle the multi-objective problem of influence spread maximization and cost minimization in social network analysis. Firstly, discrete encoding mechanism and evolutionary rules are conceived to map the original differential evolution algorithm into the topological network space. Secondly, multi-mutation strategies lead the algorithm to find uniformly distributed Pareto non-dominated solutions. Thirdly, the framework of differential evolution algorithm for the IM-CM problem is modeled. Extensive experiment results illustrate the excellent performance of the introduced algorithm in maintaining relatively large number of influence spread with the minimum cost of seed nodes. Although the proposed DMODE can accurately find the uniformly distributed Pareto non-dominated solutions, the exploiting local search strategy based on degree ranking still needs a deal of time to obtain a better solution, especially in large-scale networks. Therefore, developing effective meta-heuristic algorithms but with low time-consuming to solve IM-CM problem is one of the major of our future works.

7 Acknowledgments

This work was financially supported by the Zhejiang Provincial Natural Science Foundation under grant number LQ20F020011 and the National Natural Science Foundations of China under grant number 62162040.

References

- [1] X. Li, X. Cheng, S. Su, C. Sun, Community-based seeds selection algorithm for location aware influence maximization, *Neurocomputing* 275(2018) 1601-1613.
- [2] Q. He, X. Wang, Z. Lei, M. Huang, Y. Cai, L. Ma, Tifim: A two-stage iterative framework for influence maximization in social networks, *Applied Mathematics and Computation* 354(2019) 338-352.
- [3] S.-S. Zhang, X. Liang, Y.-D. Wei, X. Zhang, On structural features, user social behavior, and kinship discrimination in communication social networks, *IEEE Transactions on Computational Social Systems* 7(2)(2020) 425-436.
- [4] P. Wu, L. Pan, Scalable influence blocking maximization in social networks under competitive independent cascade models, *Computer Networks* 123 (2017) 38-50.
- [5] Y.-W. Teng, C.-H. Tai, P.-S. Yu, M.-S. Chen, Revenue maximization on the multi-grade product, in: *Proc. of the 2018 SIAM International Conference on Data Mining*, SIAM, 2018.
- [6] Q. He, X. Wang, M. Huang, J. Lv, L. Ma, Heuristics-based influence maximization for opinion formation in social networks, *Applied Soft Computing* 66(2018) 360-369.
- [7] X. Molinero, F. Riquelme, M. Serna, Satisfaction and power in unanimous majority influence decision models, *Electronic Notes in Discrete Mathematics* 68(2018) 197-202.
- [8] X. Molinero, F. Riquelme, M. Serna, Measuring satisfaction and power in influence based decision systems, *Knowledge-Based Systems* 174(2019) 144-159.
- [9] D. Kempe, J. Kleinberg, Maximizing the spread of influence through a social network, in: *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.
- [10] N. Sumith, B. Annappa, S. Bhattacharya, Influence maximization in large social networks: Heuristics, models and parameters, *Future Generation Computer Systems* 89(2018) 777-790.
- [11] S.S. Singh, K. Singh, A. Kumar, H. K. Shakya, B. Biswas, A survey on information diffusion models in social networks, in: *Proc. International Conference on Advanced Informatics for Computing Research*, 2018.
- [12] A. Sheikahmadi, A. Zareie, Identifying influential spreaders using multi-objective artificial bee colony optimization, *Applied Soft Computing* 94(2020) 106436.
- [13] J.F. Robles, M. Chica, O. Cordon, Evolutionary multiobjective optimization to target social network influentials in viral marketing, *Expert systems with applications* 147(2020) 113183.
- [14] J. Yang, J. Liu, Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm, *IEEE Access* 6(2017) 2320-2329.
- [15] P. Domingos, M. Richardson, Mining the network value of customers, in: *Proc. International Conference on Knowledge Discovery and Data Mining*, 2001.
- [16] M. Gong, J. Yan, B. Shen, L. Ma, Q. Cai, Influence maximization in social networks based on discrete particle swarm optimization, *Information Sciences* 367 (2016) 600-614.
- [17] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, N. Glance, Cost-effective outbreak detection in networks, in: *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007.
- [18] A. Goyal, W. Lu, L. V. Lakshmanan, Celf++: optimizing the greedy algorithm for influence maximization in social networks, in: *Proc. of the 20th International Conference Companion on World Wide Web*, 2011.
- [19] Z. Aghaee, S. Kianian, Efficient influence spread estimation for influence maximization, *Social Network Analysis and Mining* 10(1)(2020) 1-21.
- [20] G. Wu, X. Gao, G. Yan, G. Chen, Parallel greedy algorithm to multiple influence maximization in social network, *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15(3)(2021) 1-21.
- [21] M. Wang, W. Li, Y. Guo, X. Peng, Y. Li, Identifying influential spreaders in complex networks based on improved k-shell method, *Physica A: Statistical Mechanics and its Applications* 554(2020) 124229.
- [22] L. Fei, Q. Zhang, Y. Deng, Identifying influential nodes in complex networks based on the inverse-square law, *Physica A: Statistical Mechanics and its Applications* 512(2018) 1044-1059.
- [23] C. Salavati, A. Abdollahpouri, Z. Manbari, Bridgerank: A novel fast centrality measure based on local structure of the network, *Physica A: Statistical Mechanics and its Applications* 496(2018) 635-653.
- [24] J. Wang, X.-J. Ma, B.-B. Xiang, Z.-K. Bao, H.-F. Zhang, Maximizing influence in social networks by distinguishing the roles of seeds, *Physica A: Statistical Mechanics and its Applications* 604(2022) 127881.
- [25] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, K. Xie, Simulated annealing based influence maximization in social networks, in: *Proc. Twenty-fifth AAAI Conference on Artificial Intelligence*, 2011.
- [26] L. Cui, H. Hu, S. Yu, Q. Yan, Z. Ming, Z. Wen, N. Lu, Ddse: A novel evolutionary algorithm based on degree-descending search strategy for influence maximization in social networks, *Journal of Network and Computer Applications*

- 103(2018) 119-130.
- [27] Q. Ni, J. Guo, H. W. Du, H. Wang, Multi-attribute based influence maximization in social networks: Algorithms and analysis, *Theoretical Computer Science* 921(2022) 50-62.
- [28] L. Wang, L. Ma, C. Wang, N.-g. Xie, J. M. Koh, K. H. Cheong, Identifying influential spreaders in social networks through discrete moth-flame optimization, *IEEE Transactions on Evolutionary Computation* 25(6)(2021) 1091-1102.
- [29] T. K. Biswas, A. Abbasi, R. K. Chakraborty, A two-stage vikor assisted multi-operator differential evolution approach for influence maximization in social networks, *Expert Systems with Applications* 192(2022) 116342.
- [30] J. J. Lotf, M. A. Azgomi, M. R. E. Dishabi, An improved influence maximization method for social networks based on genetic algorithm, *Physica A: Statistical Mechanics and its Applications* 586(2022) 126480.
- [31] S. S. Singh, K. Singh, A. Kumar, B. Biswas, Aco-im: maximizing influence in social networks using ant colony optimization, *Soft Computing* 24(13)(2020) 10181-10203.
- [32] A. Zareie, A. Sheikhhahmadi, M. Jalili, Identification of influential users in social network using gray wolf optimization algorithm, *Expert Systems with Applications* 142(2020) 112971.
- [33] F. Qian, C. Zhu, X. Chen, J. Chen, S. Zhao, Y. Zhang, A novel approach for influence maximization based on clonal selection theory in social networks, in: *Proc. 2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018.
- [34] I. M. Ali, D. Essam, K. Kasmarik, Novel binary differential evolution algorithm for knapsack problems, *Information Sciences* 542(2021) 177-194.
- [35] X. Zhong, P. Cheng, Z. You, An improved differential evolution algorithm based on basis vector type and its application in fringe projection 3D imaging, *Knowledge-Based Systems* (2023) 110470.
- [36] F. Qian, C. Zhu, X. Chen, S. Zhao, Y. Zhang, Prnsga-ii: A novel approach for influence maximization and cost minimization based on nsga-ii, in: *Proc. 2020 IEEE International Conference on Knowledge Graph (ICKG)*, 2020.
- [37] R. Olivares, F. Munoz, F. Riquelme, A multi-objective linear threshold influence spread model solved by swarm intelligence-based methods, *Knowledge-Based Systems* 212(2021) 106623.
- [38] S. S. De, S. Dehuri, Multi-objective biogeography-based optimization for influence maximization-cost minimization in social networks, in: *Proc. International Conference on Biologically Inspired Techniques in Many-Criteria Decision Making*, 2019.
- [39] A. Sheikhhahmadi, A. Zareie, Identifying influential spreaders using multi-objective artificial bee colony optimization, *Applied Soft Computing* 94(2020) 106436.
- [40] S. Jia, L. Chen, Y. Chen, B. Li, W. Liu, Minimizing the seed set cost for influence spreading with the probabilistic guarantee, *Knowledge-Based Systems* 216(2021) 106797.
- [41] V. Roostapour, A. Neumann, F. Neumann, T. Friedrich, Pareto optimization for subset selection with dynamic cost constraints, *Artificial Intelligence* 302(2022) 103597.
- [42] E. Cunegatti, G. Iacca, D. Bucur, Large-scale multi-objective influence maximisation with network downscaling, in: *Proc. Parallel Problem Solving from Nature-PPSN XVII, PPSN 2022*, 2022.
- [43] A. Ansari, M. Dadgar, A. Hamzeh, J. Schlotterer, M. Granitzer, Competitive influence maximization: Integrating budget allocation and seed selection. <<https://arxiv.org/abs/1912.12283>>, 2019.
- [44] A. Mohammadi, M. Saraee, Finding influential users for different time bounds in social networks using multi-objective optimization, *Swarm and Evolutionary Computation* 40(2018) 158-165.
- [45] M. Pant, H. Zaheer, L. Garcia-Hernandez, A. Abraham, et al., Differential evolution: A review of more than two decades of research, *Engineering Applications of Artificial Intelligence* 90(2020) 103479.
- [46] M. Yu, W. Yang, W. Wang, G. Shen, G. Dong, L. Gong, Uggreedy: influence maximization for user group in microblogging, *Chinese Journal of Electronics* 25(2)(2016) 241-248.
- [47] S. Manchanda, A. Mittal, A. Dhawan, S. Medya, S. Ranu, A. Singh, Gcomb: Learning budget-constrained combinatorial algorithms over billion-sized graphs, *Advances in Neural Information Processing Systems* 33(2020) 20000-20011.
- [48] H. Chen, W. Qiu, H.-C. Ou, B. An, M. Tambe, Contingency-aware influence maximization: A reinforcement learning approach, in: *Proc. Uncertainty in Artificial Intelligence*, PMLR, 2021.
- [49] E. Meirom, H. Maron, S. Mannor, G. Chechik, Controlling graph dynamics with reinforcement learning and graph neural networks, in: *Proc. International Conference on Machine Learning*, PMLR, 2021.
- [50] R. Storn, K. Price, Differential evolution a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* 11(4)(1997) 341-359.
- [51] K. M. Sallam, R. K. Chakraborty, M. J. Ryan, A two-stage multi-operator differential evolution algorithm for solving resource constrained project scheduling problems, *Future Generation Computer Systems* 108(2020) 432-444.
- [52] I. M. Ali, D. Essam, K. Kasmarik, A novel design of differential evolution for solving discrete traveling salesman problems, *Swarm and Evolutionary Computation* 52(2020) 100607.
- [53] D. Bucur, G. Iacca, A. Marcelli, G. Squillero, A. Tonda, Multi-objective evolutionary algorithms for influence maximization in social networks, in: *Proc. European Conference on the Applications of Evolutionary Computation*, 2017.