2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Multi-Objective Influence Maximization Under Varying-Size Solutions and Constraints

Tarun K. Biswas* †, Alireza Abbasi*, Ripon K. Chakrabortty*
*School of Engineering and IT, University of New South Wales, Canberra 2600, Australia
†Department of IPE, Jashore University of Science and Technology, Jashore 7408, Bangladesh
(t.biswas@student.adfa.edu.au, a.abbasi@unsw.edu.au, r.chakrabortty@unsw.edu.au)

***Abstract*—Identification of a set of influential spreaders in a network, called the Influence Maximization (IM) problem, has gained much popularity due to its immense practicality. In real-life applications, not only the influence spread size, but also some other criteria such as the selection cost and the size of the seed set play an important role in selecting the optimal solution. However, majority of the existing works have treated this issue as a single-objective optimization problem, where decision-makers are forced to make their choices regarding other variables in advance despite having a thorough understanding of them. This research formulates a multi-objective version of the IM problem (referred to as MOIMP), which considers three competing objectives while subject to certain practical restrictions. Theoretical analysis reveals that the influence spreading function under the suggested MOIMP framework is no longer monotone, but submodular. We also considered three well-established multi-objective evolutionary algorithms to solve the proposed MOIMP. Since the proposed MOIMP addresses varying-size seeds, all the considered algorithms are significantly modified to fit into it. Experimental results on four real-life datasets, evaluating and comparing the performance of the considered algorithms, demonstrate the effectiveness of the proposed MOIMP.**

***Index Terms*—Multi-Objective Influence Maximization, Social Network Analysis, Evolutionary Computation, Viral Marketing.**

I. Introduction

The applications of online social networking sites like Facebook, Instagram, Twitter and so forth are now wide-spread, ranging from the spread of information/news, the promotion of commercial goods/services, and the campaign of political viewpoints [1]. Identification of few influential actors in a social network helps developing strategies towards effectively and efficiently manage those activities. In research community, the problem of identifying a subset of such influential users (called the seed set) of a social network is known as the Influence Maximization (IM) problem [2].

Most of the existing works have considered the single objective version of the IM problem focusing on the maximization of the influence spread size [3], [4]. However, in this situation, the decision regarding some important parameters such as the size of seed set, their activation/selection costs and so on are made beforehand without knowing details about them. This problem can be overcome by formulating the IM problem as a multi-objective problem and simultaneously optimizing those multiple objectives to obtain a collection of Pareto optimal solutions (regarded as Pareto Frontier/Front (PF), in which no solution is better than others across all objectives). The PF offers a broad view about the solution spaces and enables the decision makers to do trade-offs within the set.

In recent times, a few scholars [1], [5], [6] have explored the multi-objective variant of this problem. In the context of viral marketing, Robles et al., [5] formulated a bi-objective problem that considers maximization of the seed's influence spread as one objective, while minimization of their selection cost as other. In a different research, Sheikhahmadi et al. [1] looked into optimizing the distance between the seed nodes in addition to the earlier two goals. However, this third objective is co-related with the objective of maximizing the influence spread. The primary goal of increasing the distances between the seed nodes is to minimize overlapping influences, which ultimately increases the spread of total influence.

The size of the seed set (i.e., denoted as k) plays a vital role because it is linked to managerial and some other indirect costs. However, almost all the existing works have treated this as a parameter, where the value of k is needed to be fixed before solving the problem. One research group [7] highlighted this issue and proposed a bi-objective IM problem that includes minimizing k as one of its objectives. Nevertheless, they treated their proposed approach as the same way of the classical IM problem and included the seed nodes to compute the influence spread size. The consideration of seed nodes in the spread calculation may cause misjudgement in evaluating solutions when varying-size solutions are considered. For instance, when two feasible seed sets X and Y exist such that $|X| = 4$ and $|Y| = 2$, and they are, respectively, capable of activating (influencing) two and four other nodes in the network under a certain diffusion model, the classical method fails to differentiate between these two scenarios as both give a total of 6 active nodes (including the seed nodes). This is not even practical as the activation processes of seed nodes and other nodes differ. The initial seed nodes are often required to be activated or managed, which involves costs [5]. In other words, the seed nodes can be seen as investments, whereas the influenced (or activated) nodes by the seed set are achievements. Otherwise, every node in a network might be considered as a seed node in order to maximize total influence, which does not make any sense. Therefore, the initially active seed nodes should not be included in the influence spread computation, but the set of nodes activated by them.

To overcome all these concerns, we propose a new Multi-Objective Influence Maximization Problem (MOIMP) formulation considering three important objective functions: i)

IEEE/ACM ASONAM 2022, November 10-13, 2022
978-1-6654-5661-6/22/$31.00 © 2022 IEEE

maximization of the influence spread, ii) minimization of the selection cost and iii) minimization of the seed set size, k. Now, somebody can argue that there is a co-relation between second and third objective functions. This might be true for the case of equal or random cost considerations of nodes (allowing lower selection costs for higher influential users), which is often infeasible. In practice, the activation cost of a node exponentially increases(decreases) with the increase(decrease) of its influential power, particularly in the competitive environment. For example, famous film stars/sport persons are often chosen by firms to market their products because of their higher number of followers and supporters. Since the multiple companies compete in the market, only one can hire the most influential person who can spend more money. In real-life, the such person's selection cost rises/falls exponentially with the rise/fall in their influence level (i.e., the number of supporters and followers). Thus, minimizing the size of seed set does not necessarily mean to reduce the selection cost.

Due to the different characteristics, none of the current IM algorithms can tackle our suggested MOIMP. Therefore, three most well-known and powerful multi-objective evolutionary algorithms with mutually different properties from the literature of other research domains have been explored to solve it. These are: Non-dominated Sorting Genetic Algorithm II (NSGA-II) [8], Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) [9], and Non-dominated Tournament Genetic Algorithm 2 (NTGA-2) [10]. However, since the proposed MOIMP deals with varying-size solutions, incorporating them into an evolutionary optimization framework is not a trivial task. Additionally, NSGA-II and MOEA/D are originally developed based on continuous optimization problems. Therefore, significant changes are made on these algorithms; including the crossover and mutation operators, to solve the proposed MOIMP model. Finally, four real-life datasets are used to evaluate the performance of the suggested MOIMP formulation and considered algorithms. The key contributions of this work are highlighted as follows:

- We formulated the traditional IM problem as a Multi-Objective Influence Maximization Problem (MOIMP), considering three conflicting objectives, which allows decision makers to make a more informed decision.
- A detail theoretical analysis is performed, which shows that spread function is no longer monotone, but submodular, in contrast to the classical IM problem.
- A problem-specific degree based initialization technique is designed to obtain a good quality initial population.
- A modified version of crossover and mutation operators is considered in order to make the investigated algorithms fit into varying-size solutions.

The rest of the paper is organized as follows. Section 2 provides some preliminary information. Section 3 presents our proposed formulation of the MOIMP. While Section 4 describes three evolutionary algorithms for MOIMP, Section 5 discusses their experimental performances. Section 6 concludes the study by providing some future research directions.

II. Preliminaries

Generally, a Multi-objective Optimization Problem (MOP) can be mathematically formulated as follows:

$$\begin{aligned} max/min \quad & F(S) = [f_1(x), f_2(x), \dots f_m(x)] \\ \text{s.t.} \quad & x \in \Omega \end{aligned} \tag{1}$$

where, Ω refers to the decision space of a variable x and $\Omega \rightarrow R^m$ translates some real-valued objective functions (i.e.,$f_i(x), i = 1, 2, \dots, m$) which need to be optimized. The symbol m denotes the number of objectives.

In MOPs, an algorithm offers a collection of non-dominated solutions (definition II.2), rather than producing a single solution like IM, that together form a Pareto Front (PF) (definition II.3) in the feasible region of the objective space. None of the solutions in the PF are claimed to be superior than others without any further information. The decision-makers are given a clear image of the objective space by the PF, allowing them to select the best option from it.

Definition II.1 (Pareto Dominance)**.** *Given two solutions $A, B \in FS$ (feasible space), in the case of minimization, A is said to dominate B iff $\forall i = [1, 2, \dots, m]$ it holds $f_i(A) \leq f_i(B)$, and $\exists i = [1, 2, \dots, m]$ it gives $f_i(A) < f_i(B)$.*

Definition II.2 (Non-Dominated Solution)**.** *The solution $A \in FS$, in the case of minimization, is said to be non-dominated (non-inferior), iff there exists no solution $B \in FS$ that gives $f_i(B) < f_i(A)$.*

Definition II.3 (Pareto Front)**.** *The Pareto Front (PF) can be defined as $PF = \{f(A) \in FS, where \quad A \rightarrow \mathbb{R}^n \quad$ is a non-dominated solution with respect to FS$\}$.*

A. Diffusion Model

It is basically a mathematical model that captures the spreading process of the influence of users in a network. In this paper, we used the most widely-used Independent Cascade (IC) model [4]. According to the IC model, all the nodes in a candidate seed set S are assumed *active*, while others are *inactive*. Thereafter, each active node u will get single chance to activate each of its outgoing neighbor node v with a probability $p_{u,v}$. This discrete process ends when no more activation is possible. The total number of activated nodes $\sigma(S)$ at the end of the diffusion process is considered as the influence spread of the seed set S. The function is known to be monotone and submodular (definitions II.4 and II.5, respectively).

Definition II.4 (Monotone function)**.** *A function $f : 2^V \rightarrow \mathbb{R}$ is said to be monotone if for all $A \subseteq B \subseteq V$, it follows either $f(A) \leq f(B)$ or $f(A) \geq f(B)$.*

Definition II.5 (Submodular function)**.** *A function $f : 2^V \rightarrow \mathbb{R}$ is called submodular if for all $A \subseteq B \subseteq V$ and $v \in V \backslash B$, it holds $f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$. Alternatively, for any $A, B \subseteq V$, it follows $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$.*

III. Proposed Formulation of MOIMP

To make this proposed formulation of MOIMP more practical, both objective functions and constraint sets are modified from the traditional IM problems. Detailed formulation of the MOIMP is given below:

A. Objective Functions

As discussed earlier, our proposed MOIMP considered three conflicting objectives which are to be optimized concurrently. Detailed are discussed below:

1) Maximization of the influence spread: remains the key objective in our suggested framework. Unlike the traditional approaches, we only considered the nodes activated by the seed set S, not the initial activated nodes set S itself as its influence spread. It leads to an important theoretical change in the IM research. Whilst the influence spread function under the IC model is considered monotone [11], it is no longer monotone under the proposed formulation (Theorem III.1). Conventionally, a huge number of time-consuming Monte-Carlo simulations (usually more than 10,000) are used to approximate the average influence spread under the IC model. Earlier studies [12], [13] revealed that there is an exponential decay in influence probability in IC model and the global influence is limited within its 3-hop neighbor areas. Recently, Biswas et al., [14] proposed Expected Influence Score (EIS), denoted as $\sigma(S)$, to roughly estimate the influence spread size of a seed set S by considering up to its 2-hops neighbors. In this paper, we also used the EIS function (Eq. 2) as a substitute function to compute the influence spread under the IC model.

$$\begin{aligned} EIS \;&=\; \left(1+\frac{1}{|N_S^1\backslash S|}\sum_{u\in N_S^2\backslash S} p\tau_u^2\right) \\ &\quad\times \sum_{i\in N_S^1\backslash S}\left(1-(1-p)^{\tau_i^1}\right) \end{aligned} \tag{2}$$

where, N_S^1 and N_S^2 represent the seed set S's 1-hop and 2-hop neighbors and p denotes the edge influence probability and τ_i^1 and τ_u^2 refer to the number of connections of nodes i and u within N_S^1 and N_S^2, respectively.

2) Minimization of the seed cost: is our second objective function. As mentioned earlier, the seed nodes are frequently required to be managed with some benefits or financial incentives. In reality, the higher influential users, the higher their selection costs [14]. Due to the lack of knowledge, we used the node's degree centrality metrics as their local influence measures and set their costs equivalent to their centrality scores. Denoting c as the cost of node v, the selection cost C of a seed set S can be expressed as follows:

$$min \quad C(S) = \sum_{v\in S} c(v) \tag{3}$$

3) Minimization of the seed set size k: is considered as the third objective. To have greater control and administration, a fewer number of seed nodes is always preferable. Mathematically, it can be defined as follows:

$$min \quad k = |S| \tag{4}$$

B. Constraint Settings of the MOIMP

Constraints are frequently arisen in real-life optimization problems which are needed to be effectively handled by the optimization algorithm. Apart from this, finding the entire PF in the solution space is neither interested to the decision makers nor the computationally inexpensive. Therefore, to make the proposed MOIMP more pragmatic, we considered two most important real-life constraints.

1) Minimum threshold on the influence spread level: As ”maximizing the influence spread” is the key objective here, having no minimum threshold on it does not sometimes make any sense to the decision makers. To be specific, if the targeted level of influence spread coverage in a network becomes at least 50% (e.g., % of customer reach in a product marketing scenario) with the targeted seed set, solutions that belong to the PF but do not provide the required coverage, then the decision makers may lose interest in them. Eq. 5 represents the constraint where λ and $\sigma(S)$ are the required minimal level and the actual influence spread of S, respectively.

$$\sigma(S) \geq \lambda \tag{5}$$

2) Limit on the maximum size of the seed set: Similarly, the maximum size k_{max} of the seed set S should be limited, taking into account the costs of management and other indirect overhead. It can be expressed as follows:

$$|S| \leq k_{max} \tag{6}$$

Note that, the same thing may also be true for the case of the selection cost considering the total budget, however, we have not set any cost constraint here just for the sake of simplicity. However, this could also be handled in the similar fashion if it were considered.

Theorem III.1. *The influence spread function $\sigma(.)$ excluding the seed nodes under the IC model for any random seed set $S \subset V$ of a graph $G(V,E)$ is non-monotone.*

Proof. The EIS function ($\sigma(S)$) takes the 1-hop (N_S^1) and 2-hop (N_S^2) neighbor sets into account to calculate the expected influence spread of a seed set S. Here, the consideration of up to 2-hop neighbors represents a special instance of real-world problem (network with maximum length 2). Consider a node $u \in V\backslash S$ which belongs to set $N_S^1 \cup N_S^2$ and has only one degree. Now, assuming a uniform influence probability p, addition of the node u to S will cause reduction in $\sigma(S \cup \{u\})$ by $1 * p$ or by $1 * p^2$ based on its location N_S^1 or N_S^2, respectively. In contrast, if the node u belongs to set $V\backslash(S \cup N_S^1 \cup N_S^2)$ and have degree more than one, the changes in $\sigma(S \cup \{u\}) - \sigma(S)$ may be greater than or equal to zero. As the inclusion of a node to a set S does not always ensure the increase in the $\sigma(S)$ value and may encounter both drop and rise, violating the condition of monotonicity (i.e., $\sigma(S \cup \{u\}) \geq \sigma(S)$), thus, $\sigma(.)$ is not a monotone function. □

Theorem III.2. *For an arbitrary instance of the IC model, the resulting $\sigma(.)$ is a submodular function.*

Proof. Suppose, X and Y are two sets such that $X \subset Y$, and V is the union set. If N_X^1, N_Y^1, N_X^2, and N_Y^2 be the 1-hop and 2-hop neighbors sets of the X and Y respectively, then we get $(X \cup N_X^1 \cup N_X^2) \subseteq (Y \cup N_Y^1 \cup N_Y^2)$. Now, assume a node $v \notin Y$, where N_v^1 and N_v^2 are the 1-hop and 2-hop neighbor sets of v and compute the marginal gain; i.e. $\sigma(X \cup \{v\}) - \sigma(X)$. This implies the number of node in $(N_v^1 \cup N_v^2)$ which are not in $(X \cup N_X^1 \cup N_X^2)$, and it must be $\geq$ to the number of node in $(N_v^1 \cup N_v^2)$ which are not in $(Y \cup N_Y^1 \cup N_Y^2)$. More specifically, $V \backslash Y$ is the maximum size of the neighbors set (i.e., $(N_Y^1 \cup N_Y^2)$) which decreases with adding nodes to Y. Since the value of $\sigma(S)$ is equivalent to the size of the neighbor sets of S, it fulfils the requirement of submodularity $(\sigma(X \cup \{v\}) - \sigma(X) \geq \sigma(Y \cup \{v\}) - \sigma(Y))$. □

Lemma III.3. *If the seed nodes $v \in S$ are excluded from its influence spread calculation, the spread function under the traditional IC model is also non-monotone for the influence probability $p < 1$, but monotonically diminishing for $p = 1$.*

Proof. Let us assume S is the seed set for a given network $G(V, E)$. Then, $V \backslash S$ denotes the maximum possible set that S can activate. For a uniform $p = 1$, the addition of a node $u \in V \backslash S$ to S will reduce the value of $\sigma(S \cup \{v\})$ by 1 from $\sigma(S)$, meeting the requirement of monotonically decreasing $(\sigma(S \cup \{u\}) \leq \sigma(S))$. However, studies [12], [14] shows that the p is relatively low and drops down rapidly as hop-distance increases. In short, the maximum hop-area covered by the seed's influence decreases as p decreases. Therefore, based on the same idea employed in the proof of Theorem III.1, it could be demonstrated that the marginal gain $\sigma(S \cup \{u\}) - \sigma(S)$ might fluctuate based on the location and degree of u with respect to S. Thus, it is not always monotone function. □

Theorem III.4. *Finding the optimal solution under the suggested MOIMP is NP-Hard.*

Proof. The classical IM problem is previously proven NP-hard [11]. When each node's cost is assumed equal and k is fixed to any integer value, the MOIMP becomes the typical IM problem. The suggested MOIMP is likewise NP-hard since the standard IM problem is a special instance of it. □

IV. Solution Approach

As mentioned previously, we have used three well-known and powerful multi-objective evolutionary algorithms from the literature, referred as NSGA-II [8], MOEA/D [9], and NTGA-2 [10] to solve the proposed MOIMP. Although there are some strategic differences in the search and selection processes, all of them are population-based and evolve through the similar crossover, mutation and selection steps until the stopping criteria are met.

After generating a offspring population from a parent population, NSGA-II employs a non-dominated sorting process to classify the solutions of the combined population into different ranks. Thereafter, it selects solutions for the next generation based on their front rank first and then the crowding distances if they have equal rank. MOEA/D, in contrast, decomposes the entire multi-objective problem into a number of single-objective sub-problems with the help of a set of predefined weight vectors and solves each sub-problem simultaneously. Here, a scalarization function is used to evaluate the fitness of a solution. Recently, NTGA-2 is proposed primarily based on the combinatorial optimization problems which considers only the offspring population for the next generation (unlike the other multi-objective evolutionary approaches). NTGA-2 puts restrictions on the mating process (i.e., the crossover operation) to produce offspring solution where two types of selection methods: i) gap-based and ii) tournament, are alternatively used. Although we adhered to the basic framework for each of these algorithms, we modified them significantly in order to fit them into the MOIMP. Considering that the proposed MOIMP considers minimizing the size of the seed set as one of its objective functions, a novel varying-size solution representation technique was designed . Moreover, we developed a problem specific degree based population initialization technique (Algorithm 1) and a heuristic named as repair mechanism (Algorithm 2) to handle constraints. Details about these steps are described in the following subsections. Note that, although we made some changes to the algorithms under consideration, the dominant terms in the theoretical time complexity analysis remain the same, which can be found in the respective publications.

A. Degree Based Initialization

Initial guessed solutions have great impact on the search process, while randomly selected poor quality initial population may delay the convergence time. Furthermore, since a great portion of nodes in a network does not involve in the spreading process [4], motivated from the work of [3], we utilized node's degree as its local influence measure and developed a induced method of generating a good quality initial population. Algorithm 1 illustrates the steps for the proposed degree based initialization method. In line 2, all the nodes $v \in V$ are sorted in descending order based on their degree centrality scores. After a random selection of seed set size k in line 5, *upbound* is calculated in line 6 with the help of a controlling parameter ϕ, and it is gradually increased at each iteration with the increase in solution index i of population P. Once the *upbound* is calculated, the seed set is selected in line 10 by arbitrarily picking up k nodes from top-*upbound* nodes in sorted $V^{'}$. The iterations are continued until the population size NP is reached.

B. Modified Crossover and Mutation

Owing to the different characteristics of the proposed MOIMP, the original crossover and mutation operators of the considered algorithms cannot directly be applied here. Therefore, we designed a new crossover technique for the MOIMP that can effectively tackle the challenges of varying-size solutions in the search. Figure 1 illustrates an example of the proposed crossover operator for the varying-size solutions of the proposed MOIMP. While the traditional single-point crossover employs a single arbitrary crossover

Algorithm 1 Steps of degree-based population initialization

```
Input: $G$, $NP$, $\phi$, $k_{max}$;
$V \leftarrow G.nodes()$;
$V^{'} \leftarrow$ sort $v \in V$ in descending order based on G.degree(v);
$P \leftarrow \varnothing$
for $i = 1 : NP$ do
    $k \leftarrow randint(1, k_{max})$;
    $upbound \leftarrow k(i + \phi)$;
    if $upbound > |V^{'}|$ then
        $upbound \leftarrow |V^{'}|$;
    end if
    $S \leftarrow random.sample(V^{'}, upbound, k)$;
    $P.add(S)$
end for
Output: Initial population $P$ of size $NP$;
```

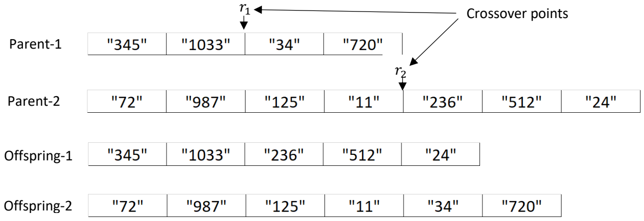

Fig. 1: An illustration of the suggested crossover operator for the MOIMP solutions with variable length

point for both solutions, the proposed crossover employs two different arbitrary cross points for two different solutions. The first random crossover point r_1 is always generated on the shortest solution (referred as parent 1 1) as follows $r_1 = randint(1, |parent1| - 1)$. Here, 1 is subtracted from the upper bound just to ensure at least one element is interchanged. In contrast, the second random crossover point r_2 is generated as $r_2 = randint(r_1, min(|parent2|, k_{max}))$. The reason behind the use of two random crossover points is that they enable the proposed crossover method to produce varying size offsprings than their parents. This is absolutely necessary in order to maintain the dynamic diversity in varying-size solutions, otherwise the final population will be identical to the initial population in terms of variation in solution lengths.

The modification in the mutation operator is not done much, except one. Generally, it iterates over each element of a crossover solution with a certain mutation probability μ to replace it with a new node from network G. In addition to this, the proposed mutation operator ensures that there is no repeating node in the offspring solution by replacing the duplicate node with 100% probability.

C. Constraint Handling Strategy

The produced offspring population might include infeasible solutions (causing the constraint violation), we designed a repair mechanism to transform them into feasible as shown in Algorithm 2. The alternatives which have less than the predefined threshold influence spreads of λ will go through the main steps from 5 to 9. An arbitrary replacement considering node's degree is repeatedly executed here until the solution S becomes feasible. More specifically, it will append a node $u \in (V - S)$ to S if $|S| < k_{max}$, else, it will replace a node $v \in S$ by a node $u \in (V - S)$ with higher degree.

Algorithm 2 Psuedocode of the repair mechanism

```
Input: $G$, $P$, $\lambda$ and $k_{max}$
1:  $V \leftarrow G.nodes()$; $P^f \leftarrow \varnothing$;
2:  for each  $S \in P$ do
3:      $spread \leftarrow \sigma(S)$ using Eq. 2;
4:      while $spread < \lambda$ do
5:          if $|S| < k_{max}$ then
6:              $S.add(v \in random(V - S))$;
7:          else
8:              replace a node $u \in S$ arbitrarily by a node $v \in (V - S)$ with higher degree;
9:          end if
10:     end while
11:     $P^f.add(S)$;
12: end for
Output: A population of feasible solutions $P^f$;
```

V. Experimental Study

The benefits of MOIMP over the existing single or two-objective problems are obvious, comparing their results makes little or no sense, especially when it comes to algorithmic performance evaluation. Because the search directions, dominance relationships, and hence the outcomes differ from one problem class to the next. Therefore, to illustrate the efficacy of the considered three algorithms in solving MOIMP, four most popular networks are used to perform our experimental study: Email [1] (nodes-1133, edges-5451) - an email correspondence network of a university; NetScience [1] (nodes-1589, edges-2742) and NetHEPT [2] (nodes-15233, edges-58891) - two different co-authorship networks; Gnutella [3] (nodes-62586, edges-147892) - a peer-to-peer file sharing network.

A. Performance indicators

Various performance indicators have been suggested over the years to evaluate the performance of an algorithm in MOPs. However, the following two most well-accepted metrics that provide both the convergence and divergence levels are employed in our study.

1) Hypervolume (HV): perhaps the most common and well-accepted performance metric in multi-objective optimization research. With a user-defined reference point (z^{ref}) (usually the worse point in the objective space), Hypervolume (HV) for a Pareto approximation set A is computed as follows:

$$HV(A) = volume\left(\bigcup_{a_i \in A} v(a_i, z^{ref})\right) \tag{7}$$

[1] http://networkrepository.com/

[2] https://arxiv.org/

[3] http://snap.stanford.edu/data/

where the function $v(a_i, z^{ref})$ represents the volume generated by these two end-points. The data of true PF is not necessary to compute the HV, which is the finest feature about HV because true PF is often unknown in reality.

2) Inverted Generational Distance Plus (IGD+): is the enhanced variant of the distance-based performance measures and recently received a lot of attention owing to its weak Pareto compliance property. IGD+ for an approximation set A is calculated with the data of actual (reference) Pareto front Z as follows:

$$IGD+(A) = \frac{1}{|Z|}\left(\sum_{i=1}^{|Z|} d_i^2\right)^{1/2} \tag{8}$$

where $d_i = max\{a_i - z_i, 0\}$, for the instance of a minimization problem, represents the Euclidean distance from $z_i \in Z$ to the nearest solution $a_i \in A$. Because the real PF for the studied problem instance is not known, motivated by the previous work [10], we considered the merged non-dominated solutions found by the investigated algorithms as the reference PF set, just to calculate the IGD+ value.

B. Parameter tuning and experimental set-up

The performance of an evolutionary algorithm greatly depends on its parameter settings. Therefore, three most important and common parameters among the three considered algorithms, referred to as population size NP, mutation probability μ and controlling parameter ϕ, are tuned on the randomly chosen *Email* dataset. We considered five different levels for each parameter and run the codes for ten times to take the average values. Here, the levels for one parameter are changed, while keeping other twos as fixed. The changes in HV and IGD+ values with the variations in parameter levels for the three parameters are given in Figures 2 and 3, respectively. Note that the higher HV and lower IGD+ values are desired for a better performance of an algorithm. The optimal parameter combinations are also reported in Table I, where almost similar trends can be noticed in both performance indices for each algorithm. Whilst the higher values for NP, the lower values for μ and ϕ are found as optimal in most cases. These optimal settings for each algorithm are used for the final comparison.

TABLE I: The optimum parameter settings for the algorithms

Performance indicator	NSGA-II			MOEA/D			NTGA-2		
	NP	μ	ϕ	NP	μ	ϕ	NP	μ	ϕ
HV	92	0.1	2	120	0.01	8	120	0.5	2
IGD+	120	0.1	2	120	0.1	2	120	0.01	2

Apart from this, some others algorithm-specific parameters are adopted from the respective papers. The widely-used Penalty Boundary Intersection (PBI) is used as the scalarization function in $MOEA/D$. The maximum seed set size k_{max} is set to 30, while influence spread is computed considering $p = 0.1$. For the sake of demonstration, the lowest level of required influence spread λ is computed by taking the top-five nodes with largest degree. As a stopping criterion, the highest number of fitness evaluation is set to 50,000 for all the algorithms. The cost of each node is computed by its degree centrality score with a arbitrary cost factor 100, just to enlarge the values and translate them into costs. Algorithmic codes are written in Python and executed in some identical Windows operated PCs with the configuration of Intel (R) Core™ i7-4770 CPU @ 3.40GHz CPU and 16GB memory.

C. Experimental results

The average values of HV and IGD+ using the non-dominated solutions obtained by different algorithms on the four datasets are given in Tables II. Here, the IGD+ values are computed on the normalized objective spaces, whereas the HVs are generated on the actual values. Considering IGD+, *NSGA-II* is clear winner among the three compared algorithms as it gives the lowest values for all the cases. However, the scenario are slightly different when it comes to HV, where $MOEA/D$ provides the highest HVs for datasets namely *NetScience* and *NetHEPT*. On the contrary, *NTGA2* fails to provide the best solution, even for a single case.

Fig. 4 displays the PFs utilizing the approximations sets produced using multiple methods on the four datasets under consideration. As observed, the obtained PF by *NSGA-II* is relatively crowded compared to the other approaches. In contrast, they are sparsely-distributed in $MOEA/D$, possibly because of its reference-based guided search strategy. It is important to note that the goals of lowering the seed set size and the cost of the seeds are in direct contrast. To be more precise, if there was a linear relationship between them, a line would emerge, not a Pareto surface. There are several non-dominated solutions spread over the three dimensional surface of the problem's objective space.

D. Runtime comparison

In this section, as an additional performance metric, we reported the computing time required by various methods, as shown in Fig. 5. Considering the runtime variations among different algorithms, the y-axis is set on a logarithmic scale to see the contrast clearly. As observed, *NSGA-II* takes the highest runtime in all cases, whereas *NTGA2* is shown as the most efficient one. Although each algorithm uses the same number of fitness evolution as the stopping criteria, differences in the selection mechanism among them create this runtime difference.

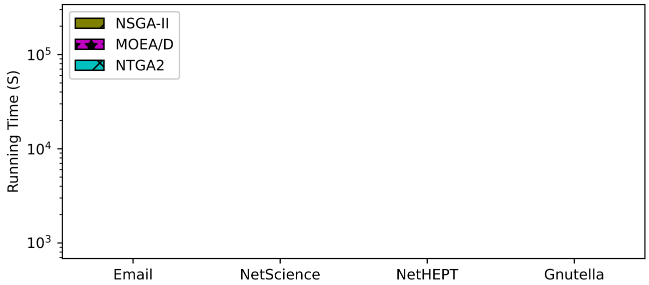

Fig. 5: Runtime comparison among different algorithms

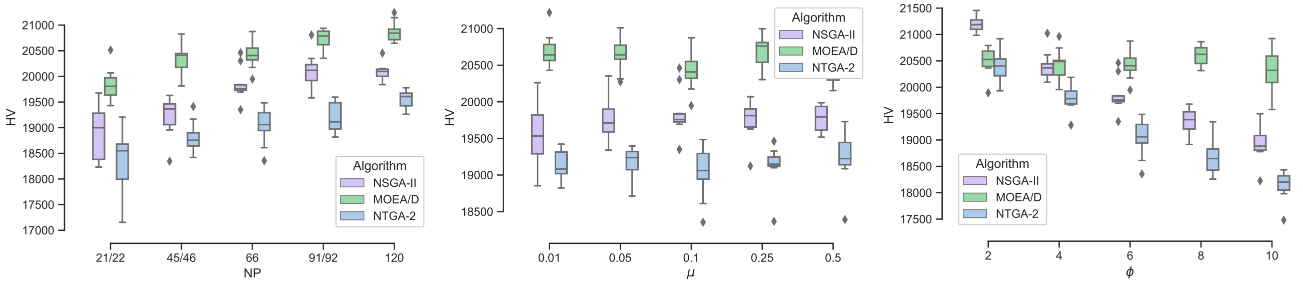

Fig. 2: The variations in HV with the changes in parameter values of the three different algorithms on *Email* network

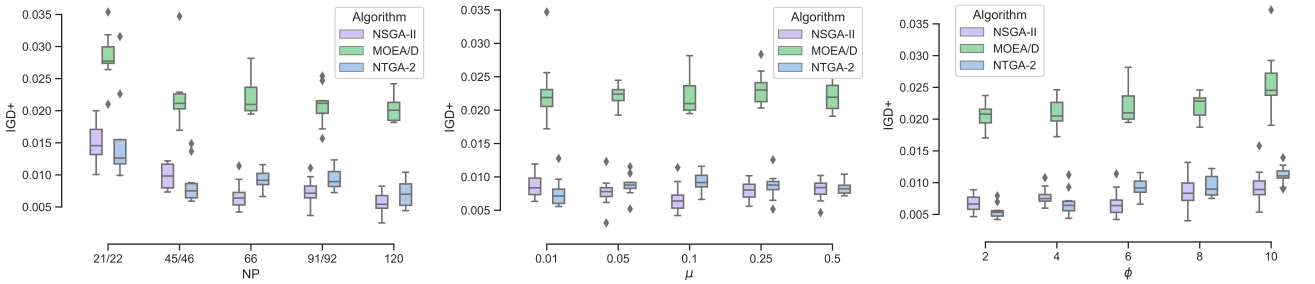

Fig. 3: The variations in IGD+ with the changes in parameter values of the three different algorithms on *Email* network

TABLE II: The calculated HV and IGD+ values for the three approaches under consideration

Metric	Algorithm	Email			NetScience			NetHEPT			Gnutella		
		Max	Mean	STD	Max	Mean	STD	Max	Mean	STD	Max	Mean	STD
HV	NSGA-II	**14054.80**	**13950.64**	86.66	4480.98	4435.39	58.46	3818.24	3796.46	21.05	**817.34**	**812.48**	4.89
	$MOEA/D$	14025.52	13924.66	100.93	**4613.42**	**4576.08**	31.70	**3856.80**	**3838.98**	13.05	784.67	784.04	0.90
	NTGA2	13593.79	13437.37	133.02	4281.81	4208.73	56.25	3701.83	3686.96	24.40	777.97	770.14	8.23
IGD+	NSGA-II	**0.0034**	**0.0045**	0.0007	**0.0056**	**0.0077**	0.0020	**0.0043**	**0.0056**	0.0011	**0.0086**	**0.0113**	0.0038
	$MOEA/D$	0.0196	0.0218	0.0016	0.0309	0.0438	0.0090	0.0332	0.0353	0.0022	0.0349	0.0367	0.0026
	NTGA2	0.0059	0.0080	0.0016	0.0105	0.0142	0.0029	0.0096	0.0120	0.0021	0.0122	0.0125	0.0003

TABLE III: Wilcoxon signed ranks test results

Dataset	NSGA-II vs MOEA/D				NSGA-II vs NTGA2				MOEA/D vs NTGA2			
	Better	Worse	P-Value	Decsision	Better	Worse	P-Value	Decsision	Better	Worse	P-Value	Decsision
Eamil	8	2	0.203	≈	10	0	0.005	+	5	5	0.142	≈
NetScience	5	5	0.203	≈	10	0	0.005	+	5	5	0.203	≈
NetHEPT	6	4	0.263	≈	8	0	0.012	+	4	6	0.214	≈
Gnutella	5	1	0.043	+	4	2	0.144	≈	3	3	0.465	≈

E. Statistical Test

Due to the stochastic character of the algorithms under consideration, we offered a statistical analysis to assess them. Firstly, we conducted Friedman test to rank the algorithms which gives the following ranking order: *NSGA-II* $>>$ $MOEA/D >>$ *NTGA2*. Although *NSGA-II* seems like the winner, we conducted another test called Wilcoxon signed ranks test to find the statistical dominance. At 5% significance level, Table III illustrates the overall test results on the four networks considering both quality measures. Here, the sign $\approx$ represents that there is no significant difference between two algorithms, while $+$ means significantly better than one another. As evident, *NSGA-II* significantly outperforms *NTGA2* on three smaller networks and *MOEA/D* on *Gnutella* dataset, while other comparisons show insignificant difference.

VI. Conclusion and Future Works

In this article, we formulated the traditional Influence Maximization (IM) problem as a Multi-Objective Influence Maximization Problem (MOIMP) which assists decision makers to make a more informed decision. Theoretical analysis shows that the influence spread function within the suggested MOIMP framework is non-monotone, yet submodular. We also considered three evolutionary algorithms from the literature to solve the problem. Since MOIMP deals with varying-size solutions, we made significant modifications in the cross-over and mutation operators to accommodate into the framework. Experimental results revealed that the proposed MOIMP is more realistic and provides better flexibility than the conventional single objective IM problem.

Exploring to more efficient algorithm for the proposed MOIMP would be an interesting future research direction. In addition, this study can be expanded by taking into account a competitive environment, where the competition is a key factor in decision-making.

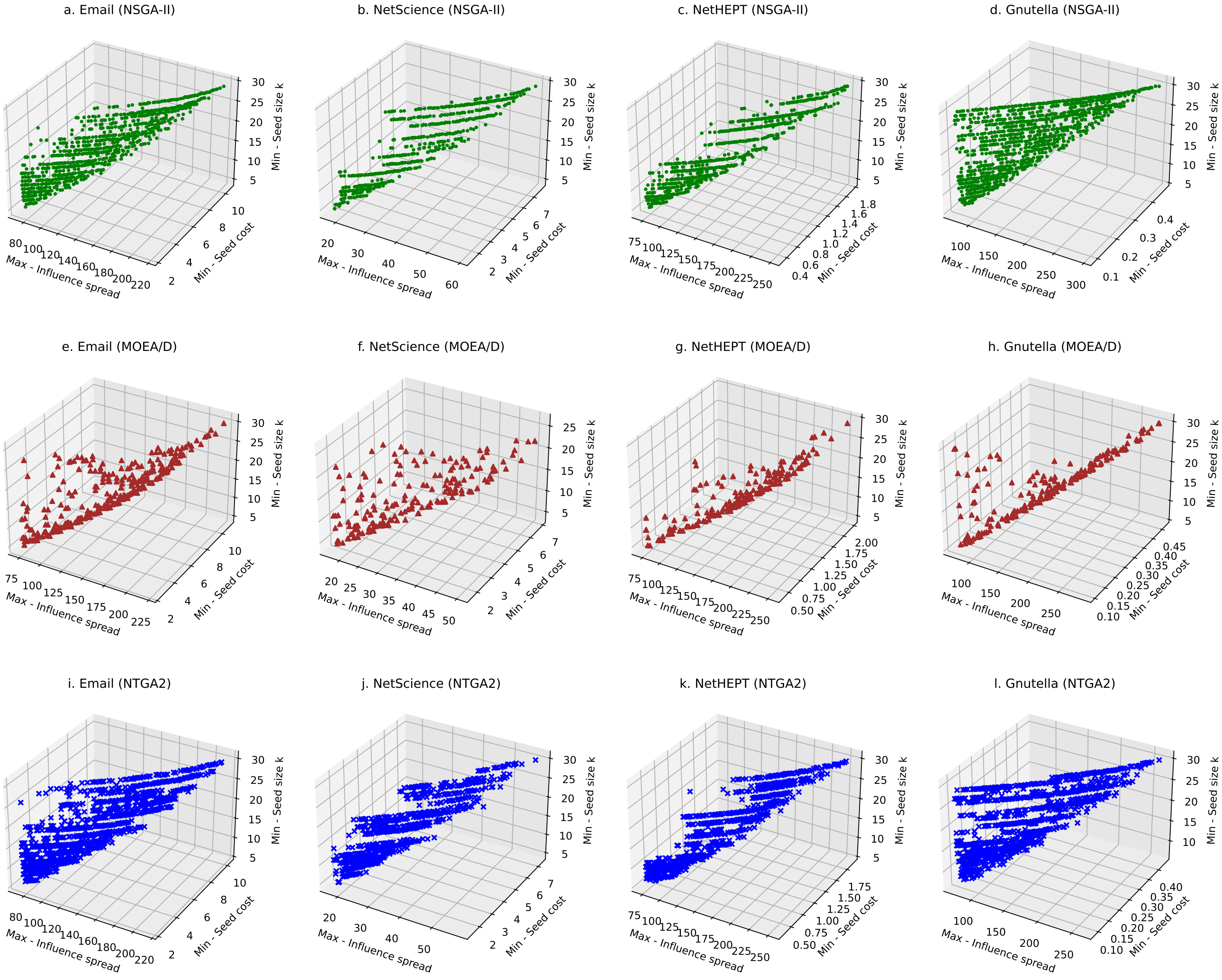

Fig. 4: The Pareto Frontier obtained by three different algorithms on the four real-world networks

References

[1] A. Sheikhahmadi and A. Zareie, "Identifying influential spreaders using multi-objective artificial bee colony optimization," *Applied Soft Computing*, vol. 94, p. 106436, 2020.

[2] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, 2015.

[3] L. Cui, H. Hu, S. Yu, Q. Yan, Z. Ming, Z. Wen, and 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*, vol. 103, pp. 119–130, 2018.

[4] T. K. Biswas, A. Abbasi, and R. K. Chakrabortty, "An mcdm integrated adaptive simulated annealing approach for influence maximization in social networks," *Information Sciences*, vol. 556, pp. 27–48, 2021.

[5] J. F. Robles, M. Chica, and O. Cordon, "Evolutionary multiobjective optimization to target social network influentials in viral marketing," *Expert Systems with Applications*, vol. 147, p. 113183, 2020.

[6] A. Mohammadi and M. Saraee, "Finding influential users for different time bounds in social networks using multi-objective optimization," *Swarm and evolutionary computation*, vol. 40, pp. 158–165, 2018.

[7] D. Bucur, G. Iacca, A. Marcelli, G. Squillero, and A. Tonda, "Multi-objective evolutionary algorithms for influence maximization in social networks," in *European conference on the applications of evolutionary computation*. Springer, 2017, pp. 221–233.

[8] 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.

[9] Q. Zhang and H. Li, "Moea/d: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on evolutionary computation*, vol. 11, no. 6, pp. 712–731, 2007.

[10] P. B. Myszkowski and M. Laszczyk, "Diversity based selection for many-objective evolutionary optimisation problems with constraints," *Information Sciences*, vol. 546, pp. 665–700, 2021.

[11] 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.

[12] M. Gong, J. Yan, B. Shen, L. Ma, and Q. Cai, "Influence maximization in social networks based on discrete particle swarm optimization," *Information Sciences*, vol. 367, pp. 600–614, 2016.

[13] N. A. Christakis and J. H. Fowler, *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown Spark, 2009.

[14] T. K. Biswas, A. Alireza, and R. K. Chakrabortty, "Selecting a cost-effective seed for maximizing the social influence under real-life constraints," *TechRxiv*, 2021, doi:10.36227/techrxiv.14489733.