

RESEARCH

Open Access



A robust optimization model for influence maximization in social networks with heterogeneous nodes

Mehrdad Agha Mohammad Ali Kermani^{1*} , Reza Ghesmati² and Mir Saman Pishvaei³

*Correspondence:

M_kermani@iust.ac.ir

¹ School of Economics, Management and Progress Engineering, Iran University of Science and Technology, Tehran, Iran

Full list of author information is available at the end of the article

Abstract

Influence maximization is the problem of trying to maximize the number of influenced nodes by selecting optimal seed nodes, given that influencing these nodes is costly. Due to the probabilistic nature of the problem, existing approaches deal with the concept of the expected number of nodes. In the current research, a scenario-based robust optimization approach is taken to finding the most influential nodes. The proposed robust optimization model maximizes the number of infected nodes in the last step of diffusion while minimizing the number of seed nodes. Nodes, however, are treated as heterogeneous with regard to their propensity to pass messages along; or as having varying activation thresholds. Experiments are performed on a real text-messaging social network. The model developed here significantly outperforms some of the well-known existing heuristic approaches which are proposed in previous works.

Keywords: Social network, Influence maximization, Influential nodes, Scenario-based stochastic programming, Robust optimization

Introduction

People often learn from each other, and this has important implications for such diverse things as how they find employment, what movies they see, which products they purchase, how technology becomes adopted, whether or not they participate in government programs or social events, and whether they protest [1–3]. The platform in which people can influence on the other's choices and decisions is social networks [4]. With the development of social network platforms in the past decade due to the growth of Internet and Web 2.0., such as Facebook, QQ, WeChat, and Micro-blog, increasing business begins to advertise their products on social networks [5–7].

One of the most important problems in social network analysis literature is influence maximization. In this problem, there exists a social agent who wants to diffuse something (such as a piece of information about advantages of a good) by way of existing social ties in a network [8]. Influence maximization is the problem of selecting a small set of seed nodes in a social network, such that their overall influence on other nodes in the network is maximized [9–11]. The selection of a minimal set of seed nodes is

constrained by high costs of exerting influence on key players. Those who would use social networks to diffuse their message seek to reach as many nodes as possible, and to do so as quickly as possible. The messages to be diffused, though, may be more effective and convincing if they are received from a friend than from the change agent, so there may be a desire to limit the number of initial contacts that are used to “seed” the diffusion [4, 12].

There are three important parameters in each diffusion process, the first parameter is the number of seed nodes in a diffusion, the second one is the total time of diffusion and the last one is the total number of nodes that are influenced in diffusion process. To leverage social influence to diffuse a message, it is desired to minimize the number of seed nodes and the total time of diffusion while maximizing the total number of infected nodes in termination of diffusion.

So, the main question which is raised in this area is, which nodes should be selected as the seed of diffusion? The existing optimization literature deals with one of the three above-mentioned parameters, but not all of them simultaneously [13]. In addition, all the previous researches which proposed a mathematical optimization model to deal with influence maximization, did not consider the probabilistic essence of the problem. They assumed that all the considered parameters of the problem are deterministic while some of them are stochastic in the real world. In addition, almost all of them assumed that the nodes are homogeneous with regard to their activation thresholds, but differing in their out-degrees (e.g., [7, 9, 12–14]), While nodes in these models may differ in the number of others to whom they have access, the previous research assumed that all nodes utilize all of their social ties. We believe that the more realistic approach is to consider nodes as heterogeneous in their propensity to act as social influencers and considering the probabilistic nature of the problem in proposed model. So, the proposed mathematical optimization model in this paper is trying to optimize two dimensions of the IM problem (the number of seed nodes and the total number of infected nodes in termination of the diffusion) simultaneously given a probabilistic influence model.

Based on [14, 15], in this paper, the node’s heterogeneity is directly measured by their “Social Skills”. It means, we believe that the better social skill of a node the more probability of forwarding a received message. So, the utilized influence model in the present paper, considers message forwarding by an “infected” node as a probabilistic process, based on their social skills.

Due to the probabilistic nature of the problem which is related to “social tie”, “mathematical models of processes on social networks”, “human behavior”, “incompleteness of observational data” and “the model parameter” [11, 16, 17], it is important to provide a solution which is capable to be robust against any realization of the probabilistic uncertainty. In the other words, the proposed solution should be immunized against uncertainty. The uncertainty of the problem has been studied in some recent works from an algorithmic point of view [16, 18, 19].

One of the well-known approaches for dealing with the mentioned uncertainties is robust optimization. Robust optimization has been proposed in the optimization literature as a modeling approach [20].

So, for the first time in this paper, a robust optimization approach is employed to develop a mathematical programming model which maximizes the expected number of

infected nodes in termination of the spread of influence and simultaneously minimize the number of seed nodes. It is worth highlighting that the main contribution of this paper in modeling and solving the influence maximization problem given a probabilistic influence model using robust optimization approach. So, the present research is proposing a robust mathematical programming model for finding the influential nodes in a certain network and while coping with the probabilistic nature of the studied problem. Based on the general advantages of robust optimization, it can be claimed that utilizing robust optimization methods may significantly enhance the efficiency of the proposed model.

So, in summary, the main contribution made by this study is proposing an integer mathematical programming model which is:

- Utilizing robust optimization approach to consider the probabilistic nature of the influence model.
- Proposing a scenario-based optimization model for influence maximization.
- Optimizing the number of seed nodes and final infected nodes simultaneously.
- Considering the heterogeneity of the nodes.

From application point of view, let a company which is deciding to diffuse a piece of information such as news on a certain social network. So, regarding the span of mobile phone in all societies [21, 22], the company selects mobile phone and particularly text messages as a tool for sending favorable information to the customer. The considered diffusion process works as follows: company sends the favorable information to some of seed nodes in network and then they will forward the short message to their friends in a probabilistic process. Therefore, the diffusion process will be occurred in some steps and then terminates when the nodes do not forward the text messages to their friends. In each step, customers who have received message are deciding to forward it to which ones of their friends.

The remainder of paper is organized as follows: “[Review of the literature](#)” section provides a brief review of recent papers which studied the influence maximization problem. The proposed optimization model and its assumptions are explained in “[Proposed optimization model](#)” section. Last section is dealing with illustrating the proposed model by implementing it on a dataset in which the nodes are students of a university and the links are their short-messages connections between them.

Review of the literature

The main problem addressed in this paper is known as “Influence Maximization”. This field of research divided to two largely separate lines of work [14]. The first deals with the competitive diffusion on networks [14, 23, 24] and the second with maximizing influence in a non-competitive situation [25, 26]. The current work falls in the second line of work. In principle, it has been proved that since the influence maximization problem could be considered as a reduced version of set covering problem, so, it is an NP-hard problem [27].

The first paper that investigated this problem from an algorithmic points of view was the work of Kempe et al. [8]. They proposed an approximation algorithm based

on a greedy strategy for finding the most influential nodes. They proved that the optimal solution can be approximated to within a factor of $(1 - \frac{1}{e} - \epsilon)$. Kempe et al. [8] took the seed nodes to be constant and optimized the number of nodes that are influenced in termination. Following their work, there are many studies which are proposing different algorithms for finding the best set of seed nodes for influence spread.

Chen and Wang [28] investigated the problem and proposed *NewGreedy* and *MixedGreedy* algorithms for finding the influential nodes in a social network. They improved the proposed algorithm by Kempe et al. [8] through reducing the running time. They evaluated their algorithms by experiments on two large academic collaboration networks obtained from the online archival database <https://arXiv.org>. Wang et al. [29] tackled influence maximization problem in a mobile phone-based social network. They noted that mobile phones are one of the most powerful tools that could be utilized in marketing, and are particularly useful in mobilizing social influence through word-of-mouth processes. They proposed a new algorithm named *Community-based Greedy Algorithm* for mining top- K influential nodes. The proposed algorithm consists of two separate parts; the first part is dealing with community detection and second part of algorithm trying to find the most influential nodes in each community. Inspiring from [29], Jalayer et al. proposed a new community-based algorithm for finding the most influential nodes in a social network. They utilized TOPSIS method as a multi attribute decision-making tool to find the influential nodes in each community [26]. Chen et al. [30] pointed out that the scalability of influence maximization is a key factor for enabling viral marketing in large scale online social networks. They developed a new heuristic algorithm that is scalable to millions of nodes. The proposed algorithm enables users to trade-off between running time and spread of influence. Another research that considered the combinatorial optimization problem of finding most influential nodes in social networks is [31]. They proposed a method of efficiently estimating the number of influenced nodes at termination based on bond percolation and graph theory; and, they provide a practical solution do the influence maximization problem on $G = (V, E)$ under the greedy hill-climbing algorithm. Wang et al. [32] investigated the influence maximization problem as the *target set selection problem*. They proposed a metaheuristic algorithm (set-based coding genetic algorithm) that converges in probability to the optimal solution of target set selection problems. They compare the results of their algorithm with the algorithm proposed by Leskovec et al. [33], the greedy algorithm developed by Kempe et al. [8], Shapley value-based influential nodes algorithm, high clustering coefficient heuristic algorithm and maximum degree heuristic algorithm.

Recently, some studies have been done in which some metaheuristic-based algorithm proposed to cope the influence maximization problem. For example, Yang and Weng [34] proposed an ant colony optimization algorithm to cope the influence maximization problem. The proposed algorithm was evaluated using a co-authorship data set and the obtained experimental results showed that the proposed algorithm outperforms two well-known benchmark heuristics. Other metaheuristic algorithm such as genetic algorithm [35], simulated annealing algorithm [36, 37], particle swarm optimization algorithm [38, 39] and cuckoo search algorithm [40] have been utilized for dealing with the influence maximization problem too. So, the researches in this

field have tried to develop approximation, heuristic or metaheuristic algorithms for finding the most influential nodes in social networks.

On the other hand, there are some recent published research in which the authors tried to use mathematical programming tools for modeling the influence maximization problem and its extensions. Kermani *et al.* developed a bi-objective integer programming model for finding most influential nodes in social network [4]. Their model dealt with minimizing the number of seed nodes and maximizing the final infected nodes simultaneously. Their research was the first one in this field which considers both cost of seeds' activation and number of final infected nodes as objectives of a mathematical programming model. Since, the considered influence model in their paper is a deterministic one, so, they solved the problem using an exact algorithm called CPLEX. They expressed that one of the simplifier assumptions in their work is considering a deterministic influence model. Following [13], there have been developed different versions of mathematical programming to tackle the influence maximization problem. For example, He *et al.* proposed a single-objective mathematical programming model to deal with the influence maximization problem [41]. They proposed a 3-hop heuristics algorithm to effectively determine the top- m influential nodes. Samadi *et al.* considered the Influence Maximization problem in presence and absence of competition. They proposed a mix-integer mathematical programming model to cope this problem [42]. Tanınmış *et al.* proposed a stochastic bilevel integer linear programming model to formulate the influence maximization. They solved the proposed model by complete enumeration for small-sized instances and by a metaheuristic for large-sized instances [43]. Guney developed a binary integer programming model for influence maximization problem. He proposed a linear programming relaxation-based method with a provable worst case bound [44]. Kermani *et al.* proposed a non-linear bi-objective mathematical programming model to tackle an extension of influence maximization problem which is named opinion aware influence maximization [15]. They proposed a genetic algorithm to solve the problem and showed its efficiency comparing with some of the state-of-the-art algorithms.

There exists another related line of research in which some related uncertainties or probabilistic nature of the information diffusion have been considered in problem modeling and solving. It seems that the study which done by He and Kempe [45] is first work that also tries to address the issue of uncertainty of parameter estimates impacting the influence maximization tasks. They investigated the problem from algorithmic point of view and did not propose any robust or non-robust mathematical programming model. Following [45], He and Kempe investigated the concept of stability in influence maximization problem when it is dealing with noise and uncertainty [17]. Chen *et al.* proposed a new problem in which the goal is to find the best possible seed set for influence maximization purpose, while considering the adverse effect of the uncertainty. They utilized the robust optimization concepts and used the worse-case multiplicative ratio between the influence spread of the chosen seed set and the optimal seed set as their objective function. It should be noted that they did not propose a mathematical programming model in their research. In another published research, Kalimeris *et al.* [46] worked on the issue of robust influence maximization in hyperparametric models. The main question they addressed in their research is whether there is a computationally efficient algorithm to

perform robust optimization for hyperparametric models or not? They worked on finding the related algorithm and proving its efficiency. However, they did not model the influence maximization problem using robust optimization mathematical tools. Based on the applying mathematical modeling, the closest work to the present research is [47]. The authors defined the general two-stage stochastic submodular optimization model and applied it to model the influence maximization problem. Then, they utilized a delayed constraint generation algorithm to find the optimum solutions. It should be noted that they did not model the considered influence model as constraints of their model. In addition, in the present work, we utilized scenario-based programming to cope the existing uncertainty which has not been done in [47].

There is not, however, any robust optimization model for modeling the maximization of the spread of information and minimization the size of seed nodes set simultaneously with an exact solution. So, a novelty of the present work is dealing with the above-mentioned objective with considering a probabilistic influence model. The other novelties of the present work are considering the heterogeneity of the nodes and the probabilistic nature of the problem in a robust optimization model simultaneously.

In addition, almost all the previous works (except [4, 15]) on the diffusion problem have focused on locating the optimal (fixed number of) nodes to maximize diffusion without considering the cost of seed nodes' activation. In many contexts, however, efforts to leverage social influence to maximize diffusion in existing social networks are costly. Those who would diffuse their message may need to provide incentives to seed nodes, or invest heavily in education and influence of initial targets in order to start the process. Our model assumes that the optimal choice of seed nodes must minimize these costs simultaneously with seeking maximal diffusion of the message.

Proposed optimization model

Let us focus on a company which decides to advertise its good or service using viral marketing; that is, influencing a small number of actors directly, and utilizing these nodes to spread the message through their social networks. One common medium for such a marketing campaign is a short-message-system (e.g., texting). Since sending a text message is costly, and costs rise directly with the number of contacts that are initially made. In addition, the sending of more messages directly from the company, the less forwarding may occur, and messages may be fewer effective influencers because they have not been forwarded within existing relationships of trust among friends. Consequently, it is in the interest of the company to minimize the number of initial contacts. At the same time, the main goal of the company seeing that the text message reaches the largest number of members of the target population. So, the company would like to target seed nodes that have many social ties, and who are willing to pass along the message.

Considered diffusion model

The considered message passing process (diffusion model) in the present work which is exactly similar to the considered model in [4], is as follows:

Let a network $G = (V, E)$, where V and E is the set of nodes and links, respectively.

- Persons or nodes (V) are embedded in a social network (G), and may receive communications from, and communicate to, discrete numbers of other individuals (E). Connections (links) are directional, and may be reciprocated ($G = (V, E)$).
- Message (information) diffusion occurs along existing social networks, and is stochastic. That is, activated nodes ($i \in V$) may, or may not forward messages with a fixed probability. In the other words, the considered diffusion model is a stochastic one.
- The probability that a person forwards a message, is directly proportional to their sociability. Persons with more social skills are more likely to forward a message, regardless of their out-degree.
- Each person (node) has either received a message (is activated), or has not (is inactive). Once activated, a node remains activated. In the other words, the considered influence model is a progressive one.
- Time is treated as discrete intervals during which forwarding by activated nodes can occur.
- Activated nodes may forward a message only within one time period of receiving it.

Message diffusion occurs as a probabilistic process, based on social ties' propensity to act as social influencers. In the other words, person i forwards a piece of information to person j with the probability of p_{ij} . Based on [4], this probability can be obtained through $\frac{p_i \cdot p_j}{\sum_{j \in N_i} p_i \cdot p_j}$, in which p_i is the probability of forwarding message by i . Furthermore, p_i is estimated using the social skill questionnaire score of person i [4], that is F_i ; a simple way to estimate p_i may be $p_i = \frac{F_i}{\max_i F_i}$. The probabilistic essence of the considered diffusion model is modeled by p_i . It should be noted that the considered probabilistic diffusion model is as most as possible accordance with the real-world message passing through mobile phones. The considered assumptions in considered diffusion model in the present research are different form classical diffusion models such as Linear Threshold (LT) and Independent Cascade (IC). For example, in LT diffusion model, each link has a certain and predefined weight which has a key role in activation regime. In addition, each node has a randomly predefined sensitive threshold for being activated. But in the considered diffusion model in the present paper, the nodes have no sensitive threshold and could be activated based on a probability. On the other hand, in IC, each newly activate node ($i \in V$) has a single chance of activating each of its inactive out-neighbors ($j \in V$) with probability p_{ij} . So, the considered diffusion model in this paper can be considered as an extension of IC, in which, the p_{ij} is proportional to the social skill of source and sink nodes.

Notation

To cope with the probabilistic nature of the problem, a robust scenario-based stochastic programming model is developed. Each scenario in this model specifies a set of potentially activated links between the nodes which may be generated randomly based on p_{ij} . It should be noted that actual activation of links in each scenario is related to three factors:

- The seed nodes which are independent of scenarios.

- Links potentially activated in each scenario.
- Nodes activated in different time periods, except the initial time, in each scenario.

The notation that is used to propose the robust optimization model (ROM) is shown in Table 1.

a_{ijs} is the parameter that defines different scenarios based on p_{ij} . It determines if a message is received by the person i at a time period whether he forwards the message to the person j ($j \in N_i$) in scenario s . It should be noted that in this model x_i^0 is the only decision variable which can be determined by the social change agent. Furthermore, this variable is independent of scenarios as a first stage variable.

Scenario-based stochastic influence maximization problem

In terms of the expressed notations, the scenario-based stochastic influence maximization model can be formulated as follows:

$$\text{Min} \sum_{i=1}^n x_i^0 \tag{1}$$

$$\text{Max} \sum_{s=1}^S \pi_s Z_s \tag{2}$$

s.t.

Table 1 The notations which are used to formulate the problem

Sets and indices	
i, j	Index of source and destination persons $i, j = 1, \dots, n$;
t	Index of discrete time periods $t = 0, 1, \dots, T$;
s	Index of scenarios $s = 1, \dots, S$;
N_i	Out-degree of person i (i.e., the persons that i has their phone number in her contact list);
K_i	In-degree of person i (i.e., the persons that have i 's phone number in their contact list)
Parameters	
a_{ijs}	Binary parameter representing that the message may be forwarded by the person i to the person j in scenario s ;
π_s	Probability of scenario s ;
M	An adequate large number
Variables	
x_i^0	$= \begin{cases} 1 & \text{if the person } i \text{ received a message at the initial time of the message diffusion,} \\ 0 & \text{otherwise;} \end{cases}$
x_{is}^t	$= \begin{cases} 1 & \text{if the person } i \text{ received a message at time period } t \text{ in scenario } s (t = 1, \dots, T), \\ 0 & \text{otherwise;} \end{cases}$
l_{ijs}^t	$= \begin{cases} 1 & \text{if the person } i \text{ forward the message to the person } j \text{ at period } t \text{ in scenario } s, \\ 0 & \text{otherwise;} \end{cases}$

$$Z_s = \sum_{i=1}^n x_{is}^T, \quad \forall s, \tag{3}$$

$$l_{ijs}^0 \leq a_{ijs} x_i^0, \quad \forall i, j \in N_i, s, \tag{4}$$

$$l_{ijs}^t \leq a_{ijs} x_{is}^t, \quad \forall i, j \in N_i, s, t = 1, \dots, T, \tag{5}$$

$$\sum_{i \in K_j} a_{ijs} l_{ijs}^t \leq M x_{js}^{t+1}, \quad \forall j, s, t = 0, \dots, T - 1, \tag{6}$$

$$\sum_{i \in K_j} a_{ijs} l_{ijs}^t \geq (x_{js}^{t+1} - x_{js}^t), \quad \forall j, s, t = 0, \dots, T - 1, \tag{7}$$

$$x_i^0 \leq x_{is}^1, \quad \forall i, s, \tag{8}$$

$$x_{is}^t \leq x_{is}^{t+1}, \quad \forall i, s, t = 1, \dots, T - 1, \tag{9}$$

$$\sum_{j \in N_i} l_{ijs}^1 \leq M (x_{is}^1 - x_i^0), \quad \forall i, s, \tag{10}$$

$$\sum_{j \in N_i} l_{ijs}^{t+1} \leq M (x_{is}^{t+1} - x_{is}^t), \quad \forall i, s, t = 1, \dots, T - 1, \tag{11}$$

$$x_i^0 \in \{0, 1\}, \quad \forall i, \tag{12}$$

$$x_{is}^t \in \{0, 1\}, \quad \forall i, s, t = 1, \dots, T - 1, \tag{13}$$

$$l_{ijs}^t \in \{0, 1\}, \quad \forall i, s, t. \tag{14}$$

The model seeks an optimum of maximizing the number of nodes reached by the message in a fixed period of time, while remaining sensitive to minimizing costs of influencing “key players”. Objective function (1) is related to minimizing the number (and hence cost) of nodes that are initially activated. The objective function (2) is associated with maximizing the expected number of activated nodes at the end of a fixed period. Z_s in the Objective function (2) is obtained from Eq. (3). Constraints (4) and (5) assure that if a link is active at t in scenario s , then its source node is also active. If a node is inactive at t in scenario s , then its outgoing links are inactive. Further, these constraints show that if a node is active at t in scenario s , its outgoing links could be active or inactive. Constraint (6) states that if a link is active at t in scenario s , then the destination node is active at $t + 1$ in scenario s . Furthermore, a node is inactive at $t + 1$ if and only if all the incoming links are inactive at t . Constraint (7)

indicates that if a node is active at $t + 1$ and inactive at t in scenario s , then at least one of the incoming links should be active at the former time in the same scenario; as well if a node is active at both t and $t + 1$ in scenario s , then the incoming links may be active or inactive at t in scenario s . In some previous works [8] node activation is based on independent cascade or linear threshold logics. Since the proposed model is dealing with diffusion through short message systems, the influence process should be modeled according to reality of SMS diffusion. In reality when a short message is received by mobile phone, we read it and will be active. Constraints (8) and (9) are included to make the second objective true. These constraints try to make all the nodes that are active in each stage also active at last stage. Constraints (10) and (11) indicate that if a node is active or inactive at both t and $t + 1$ in scenario s , its outgoing links should become inactive. These constraints prevent against unreasonable activation of links by limiting the period of time that they can activate others. That is, nodes activate others for a limited period of time after their own situation changes. Parameter M in Constraints (7), (10), and (11) is a reasonably large number. Finally, Eqs. (12)–(14) show the type of decision variables. Notably, above system constraints should be satisfied in all scenarios.

The proposed robust optimization model

The philosophy of robust programming is based on risk-averse methods to conserve the optimal solution for any realization of uncertain parameters. A solution to an optimization problem is said to be robust if it has both “feasibility robustness” and “optimality robustness”. Feasibility robustness indicates that the solution should stay feasible for almost all plausible values of uncertain parameters and optimality robustness means that the objective function value for the solution should stay near to optimal value or have minimum deviation from the optimal value for almost all plausible values of uncertain parameters [48].

Soyster played a pioneering role in developing the robust optimization theory [49]. He presented a worst-case robust programming method for inexact linear programming problems. Thereafter robust optimization approach has developed in three lines: (i) robust scenario-based stochastic programming [50]. (ii) Robust programming based on closed convex uncertainty sets [51–55] (iii) Robust possibilistic programming [48].

Mulvey et al. introduced a robust optimization approach for scenario-based stochastic programming models by presenting a trade-off between optimality robustness and feasibility robustness (which is called “solution robustness” and “model robustness”, respectively, in their work) [50]. The optimality robustness is modeled by adding a weighted variability measure of objective function of scenarios to the expected value of them. Varying the weight put on this variability drives the optimization process to provide solutions that may present higher expected total costs with lower cost-deviations under different scenarios. Several measures are developed to specify the variability of scenarios. Mulvey et al. recommend the variance of scenarios objective function [50]. Due to the non-linear form of the variance function [56, 57], have attempt to convert the problem into a linear programming model.

Due to the probabilistic nature of the presented problem in this paper, the model should be robust against any realization of stochastic scenarios, meaning that the

proposed solution should have the least variability under different scenarios. Here, we have used the proposed approach in [57] to develop the robust stochastic counterpart of the proposed model which is provided as follows:

$$\text{Min} \sum_{i=1}^n x_i^0 \tag{1}$$

$$\text{Max} \sum_{s=1}^S \pi_s Z_s - \lambda \sum_{s=1}^S \pi_s \left(Z_s - \sum_{s'=1}^S \pi_{s'} Z_{s'} + 2u_s \right) \tag{15}$$

s.t.

(3)–(14);

$$Z_s - \sum_{s'=1}^S \pi_{s'} Z_{s'} + u_s \geq 0, \quad \forall s, \tag{16}$$

$$u_s \geq 0, \quad \forall s. \tag{17}$$

Objective function (15) is the developed version of objective function (2). The second term of (15), along with constraint (16), relates to minimizing the variability of scenarios which is identified by the variability measure presented by Leung et al. [57]. This term controls optimality robustness of the model. λ is a parameter which determines the importance degree of optimality robustness in comparison with the expected number of activated nodes in the last period. Furthermore, u_s is the variable used to convert the primary non-linear problem into its equivalent linear form.

Single-objective counterpart of the model

The proposed robust optimization model is a bi-objective mixed integer linear programming which its conflicted objectives are “minimizing the cost (number of seed nodes)” and “maximizing the number of influenced nodes”. To cope with the multiple objectives nature of the proposed models, the common use ε -constraint method [58] is utilized. This approach has been used in a similar study which is done in 2016 [4]. The equivalent single-objective model is presented as follows:

$$\text{Max} \sum_{s=1}^S \pi_s Z_s - \lambda \sum_{s=1}^S \pi_s \left(Z_s - \sum_{s'=1}^S \pi_{s'} Z_{s'} + 2u_s \right) \tag{15}$$

s.t.

(3)–(14), (16)–(17);

$$\sum_{i=1}^n x_i^0 \leq \varepsilon. \tag{18}$$

Noteworthy, since ε can hold integer numbers, its intuitive interpretation is the number of seed nodes.

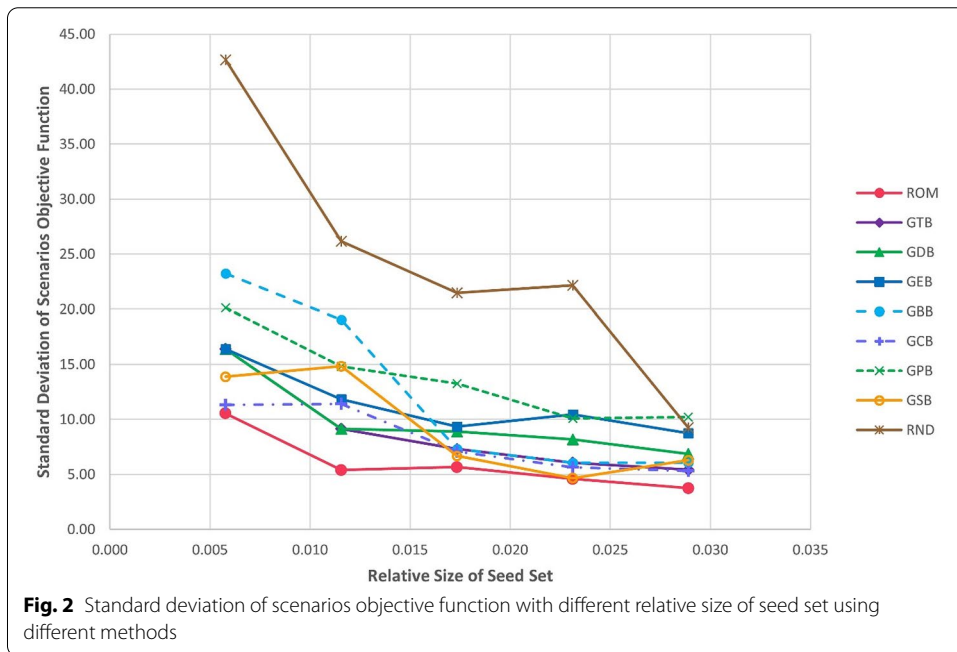
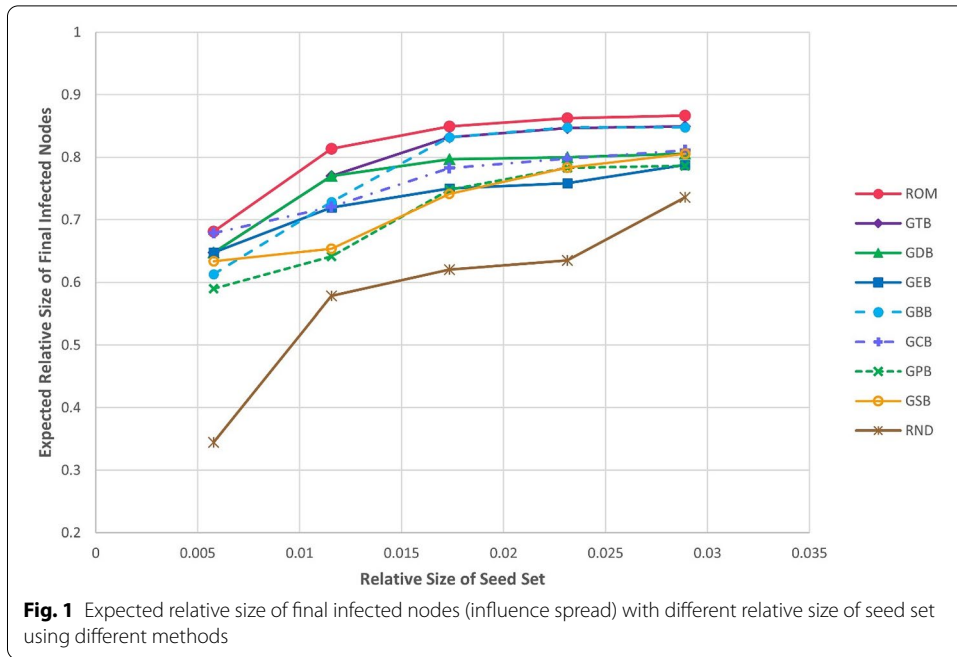
Table 2 Results of proposed model in comparison with some heuristics with activation of three seed nodes

Methods	RND	GDB	GEB	GBB	GCB	GPB	GSB	GTB	ROM
Influential nodes	41, 117, 120	30, 32, 85	17, 30, 32	161, 85, 30	173, 162, 137	85, 81, 142	32, 43, 157	30, 85, 161	74, 85, 161
Number of final infected nodes									
Scenario 1	123	138	121	151	143	137	129	151	152
Scenario 2	121	130	121	142	122	130	128	142	146
Scenario 3	102	135	129	136	137	137	121	136	147
Scenario 4	79	142	132	151	148	139	133	151	151
Scenario 5	71	127	127	134	133	116	125	134	133
Scenario 6	112	144	134	147	136	124	126	147	146
Scenario 7	104	151	146	150	129	134	140	150	150
Scenario 8	141	145	136	151	137	149	123	151	153
Scenario 9	124	143	137	144	134	126	137	144	146
Scenario 10	96	123	114	133	135	102	120	133	145
Average	107.3	137.8	129.7	143.9	135.4	129.4	128.2	143.9	146.9
Standard deviation	21.49	8.90	9.36	7.31	7.07	13.27	6.68	7.31	5.67

Case study implementation and evaluation

To illustrate the utility of the model in identifying the best seed nodes of a social network for maximizing the diffusion of information, the *Abrar* dataset [59, 60] is utilized. During 2010–2011, 163 students in two disciplines at Abrar University (Industrial Engineering and Software Engineering) were interviewed. Each of the students was asked to identify the other students who were in their mobile phone contact list. These contacts identify a directed tie from each student to others. To assess the propensity or willingness to contact others, each student also filled out a *Social Skill* questionnaire that indicates their willingness to contact others [61]. The questionnaire has 40 items grouped into two scales, Prosocial Behavior, which assesses cooperative, helping, and friendly behaviors (for example, “*I offer my classmates help to do their homework*”) and Antisocial Behavior, which assesses aggressive behaviors, disruptive reactions, and attention seeking (for example, “*I hit other kids when they make me mad*”). The items are rated on a 6-point Likert scale ranging from 1 (*it doesn't describe me at all*) to 6 (*it describes me completely*). So, a high score on the index means that a person's scores high on the pro-social, and low on the anti-social items. The probability of forwarding message from each student to others is calculated based on the *Social Skill* questionnaire and then 10 scenarios are generated randomly based on this probability. It is assumed that the probability of each scenario is equal to 0.1.

Results of implementing the proposed robust optimization model (ROM) in the *Abrar* dataset (which is used in [4, 60, 62]), and its comparison to some of the existing heuristic algorithms are shown in Table 2, Figs. 1, 2. Notably, all the results are obtained by CPLEX solver of GAMS optimization software on a Core i7 computer with 8.0 GB RAM in 2.1 s. In CPLEX, an optimality parameter can be specified to decide whether to find the optimal solution or to quickly obtain a suboptimal solution [63]. Because CPLEX



uses branch-and-cut algorithm when solving integer linear programming model, the optimal solutions can be found by setting the possible gap equal to zero. Many studies have used obtained results through running it as the benchmark solutions [13, 64]; reasonably, the performance and optimality of the obtained results have been proved. Furthermore, as all the previous works used heuristic or approximation algorithm for finding the optimal solution, it is a trivial fact that the obtained solution in this research

is better than the other research. Inspiring from [8, 14, 26, 65], the alternative heuristic algorithms for finding the most influential nodes are the *Greedy Degree Based (GDB)*; a simple heuristic that selects the k nodes with the largest degrees [3], *Greedy Eigenvector Based (GEB)*; a simple heuristic that selects the k nodes with the largest eigenvector. GEB is suggested as a heuristic algorithm in [66], *Greedy Betweenness Based (GBB)*; a simple heuristic that selects the k nodes with the largest Betweenness, *Greedy Closeness Based (GCB)*; a simple heuristic that selects the k nodes with the largest Closeness, *Greedy Pagerank Based (GPB)*; a simple heuristic that selects the k nodes with the largest Pagerank, *Greedy Topsis Based (GTB)*; selecting the k nodes with the largest Topsis scores (this ranking method is proposed and used in [15, 60, 67–69]), *Greedy Sociability Based (GSB)*; Beside the existing simple method, the other simple heuristic can be selection of the k nodes with the largest social skill which is extracted by Social Skill questionnaire [61], and finally *Random method (RND)*; simply select k random nodes in the graph.

Table 2 shows the results of the most influential nodes, number of final infected nodes in each scenario, average and standard deviation of final infected nodes using proposed ROM and mentioned heuristics. As can be seen, not only the average final infected nodes of scenarios from ROM is substantially better than other methods but also almost all scenarios have better performance in infecting nodes in final time period.

The results depicted in Fig. 1 show that among the considered methods, the ROM has the highest expected number of final infected nodes for all different numbers of seed nodes. It should be noted that despite other heuristic methods, the solution of ROM, i.e., the most influential nodes, is a global optimized solution. Figure 2 demonstrates that the ROM has the smallest standard deviation of influence spreads in different scenarios, which shows the greater robustness of the proposed ROM compared to the others. For all methods, including ROM, increasing the number of seed nodes increases the expected number of final infected at decreasing rates. Further, increasing the number of seed nodes decreases the standard deviation of final infected nodes, or increases robustness. This issue reflects the multi-objective nature of the problem. The desired solution can be determined by the social agent by making a trade-off between the two objectives, which are the number of seed nodes and the resulting costs and the expected number of final infected nodes.

Conclusions

Influence maximization is the problem of finding most influential nodes in a network to maximize the spread of influence. The proposed model outperforms plausible alternative approaches to the influence maximization/cost minimization problem on fixed social networks where the probabilistic nature of the problem originates from heterogeneity in social actors propensity to act as social influencer. So, in this paper a multi-objective robust stochastic programming model is developed which optimizes the diffusion and minimizes the number of seed nodes as a costly activity simultaneously. The model is implemented by using a real data set and the achieved results demonstrate significant increases in the expected number of final infected nodes as well as robustness of the solution in comparison with some common heuristic algorithm. Developing the

proposed ROM to a model which is capable to optimize the time of diffusion can be considered as an important direction for the future research.

Acknowledgements

Not applicable.

Authors' contributions

MAMAK and MP designed the research. RG performed the experiments and wrote some part of manuscript. The other part of manuscript has been written by MAMAK. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The dataset which is analyzed in this research is published in [59].

Declarations

Competing interests

The authors declare that they have no competing interest.

Author details

¹School of Economics, Management and Progress Engineering, Iran University of Science and Technology, Tehran, Iran.

²School of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran. ³School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran.

Received: 31 May 2020 Accepted: 26 July 2021

Published online: 27 August 2021

References

1. Jackson, M.O.: *Social and Economic Networks*. Princeton University Press, Princeton (2010)
2. Kross, E., Chandhok, S.: How do online social networks influence people's emotional lives? In: *Sydney Symposium of Social Psychology, Applications of Social Psychology, 2020*
3. Kermani, M.A.M.A., Sani, S.A., Zand, H.: Resident's Alzheimer disease and social networks within a nursing home. In: *International Conference on Complex Networks and their Applications, Springer (2020)*
4. Agha Mohammad Ali Kermani, M., Aliahmadi, A., Hanneman, R.: Optimizing the choice of influential nodes for diffusion on a social network. *Int. J. Commun. Syst.* **29**, 1235–1250 (2015)
5. Lu, F., et al.: Scalable influence maximization under independent cascade model. *J. Netw. Comput. Appl.* **86**, 15–23 (2017)
6. Bindu, P.V., Thilagam, P.S.: Mining social networks for anomalies: methods and challenges. *J. Netw. Comput. Appl.* **68**, 213–229 (2016)
7. Hegeman, J., et al.: Sponsored advertisement ranking and pricing in a social networking system, Google Patents (2020)
8. Kempe, D., Kleinberg, J., Tardos, É.: 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)*
9. Zhao, J., et al.: Competitive seeds-selection in complex networks. *Physica A Stat. Mech. Appl.* **467**, 240–248 (2017)
10. Wang, Y., et al.: Real-time influence maximization on dynamic social streams. *Proc. VLDB Endow.* **10**(7), 805–816 (2017)
11. Ju, W., et al.: A new algorithm for positive influence maximization in signed networks. *Inf. Sci.* **512**, 1571–1591 (2020)
12. Yan, Q., et al.: Group-level influence maximization with budget constraint. In: *International Conference on Database Systems for Advanced Applications, Springer (2017)*
13. Agha Mohammad Ali Kermani, M., Aliahmadi, A., Hanneman, R.: Optimizing the choice of influential nodes for diffusion on a social network. *Int. J. Commun. Syst.* **29**(7), 1235–1250 (2016)
14. Kermani, M.A.M.A., et al.: A novel game theoretic approach for modeling competitive information diffusion in social networks with heterogeneous nodes. *Physica A Stat. Mech. Appl.* **466**, 570–582 (2017)
15. Kermani, M.A.M.A., Ghesmati, R., Jalayer, M.: Opinion-aware influence maximization: how to maximize a favorite opinion in a social network? *Adv. Complex Syst.* **21**(06n07), 1850022 (2018)
16. He, X., Kempe, D.: Robust influence maximization. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 885–894. ACM, San Francisco (2016)
17. He, X., Kempe, D.: Stability and robustness in influence maximization. *ACM Trans. Knowl. Discov. Data (TKDD)* **12**(6), 1–34 (2018)
18. Chen, W., et al.: Robust influence maximization. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 795–804. ACM, San Francisco (2016)
19. Jung, K., Heo, W., Chen, W.: Irie: scalable and robust influence maximization in social networks. In: *Data Mining (ICDM), 2012 IEEE 12th International Conference on, IEEE (2012)*
20. Marotta, A., et al.: A fast robust optimization-based heuristic for the deployment of green virtual network functions. *J. Netw. Comput. Appl.* **95**, 42–53 (2017)
21. Arminen, I.: Mobile communication society? *Acta Sociol.* **50**, 431–437 (2007)

22. Campbell, S.W., Russo, T.C.: The social construction of mobile telephony: an application of the social influence model to perceptions and uses of mobile phones within personal communication networks. *Commun. Monogr.* **70**(4), 317–334 (2003)
23. Alon, N., et al.: A note on competitive diffusion through social networks. *Inf. Process. Lett.* **110**(6), 221–225 (2010)
24. Small, L., Mason, O.: Nash Equilibria for competitive information diffusion on trees. *Inf. Process. Lett.* **113**(7), 217–219 (2013)
25. Shang, J., et al.: CoFIM: a community-based framework for influence maximization on large-scale networks. *Knowl.-Based Syst.* **117**, 88–100 (2017)
26. Jalayer, M., Azheian, M., Kermani, M.A.M.A.: A hybrid algorithm based on community detection and multi attribute decision making for influence maximization. *Comput. Ind. Eng.* **120**, 234–250 (2018)
27. Lu, Z., et al.: The complexity of influence maximization problem in the deterministic linear threshold model. *J. Comb. Optim.* **24**(3), 374–378 (2012)
28. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2009)
29. Wang, Y., et al.: Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2010)
30. Chen, W., Wang, C., Wang, Y.: Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2010)
31. Kimura, M., et al.: Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Discov.* **20**(1), 70–97 (2010)
32. Wang, C., et al.: A global optimization algorithm for target set selection problems. *Inf. Sci.* **267**, 101–118 (2013)
33. Leskovec, J., et al.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2007)
34. Yang, W.-S., et al.: Application of the ant colony optimization algorithm to the influence-maximization problem. *Int. J. Swarm Intell. Evol. Comput.* **1**(1), 1–8 (2012)
35. Bucur, D., Iacca, G.: Influence maximization in social networks with genetic algorithms. In: *EvoApplications*, No 1 (2016)
36. Jiang, Q., et al.: Simulated annealing based influence maximization in social networks. In: *AAAI* (2011)
37. Liu, S.-J., Chen, C.-Y., Tsai, C.-W.: An effective simulated annealing for influence maximization problem of online social networks. *Procedia Comput. Sci.* **113**, 478–483 (2017)
38. Gong, M., et al.: Influence maximization in social networks based on discrete particle swarm optimization. *Inf. Sci.* **367**, 600–614 (2016)
39. Tang, J., et al.: Identification of top-k influential nodes based on enhanced discrete particle swarm optimization for influence maximization. *Physica A Stat. Mech. Appl.* **513**, 477–496 (2019)
40. Gandomi, A.H., Yang, X.-S., Alavi, A.H.: Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. *Eng. Comput.* **29**, 17–35 (2013)
41. He, Q., et al.: Heuristics-based influence maximization for opinion formation in social networks. *Appl. Soft Comput.* **66**, 360–369 (2018)
42. Samadi, M., et al.: Seed activation scheduling for influence maximization in social networks. *Omega* **77**, 96–114 (2018)
43. Tanınmış, K., Aras, N., Altinel, I.K.: Influence maximization with deactivation in social networks. *Eur. J. Oper. Res.* **278**(1), 105–119 (2019)
44. Güney, E.: An efficient linear programming based method for the influence maximization problem in social networks. *Inf. Sci.* **503**, 589–605 (2019)
45. He, X., Kempe, D.: Stability of influence maximization. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2014)
46. Kalimeris, D., Kaplun, G., Singer, Y.: Robust influence maximization for hyperparametric models. *arXiv preprint arXiv:1903.03746* (2019)
47. Wu, H.-H., Küçükyavuz, S.: A two-stage stochastic programming approach for influence maximization in social networks. *Comput. Optim. Appl.* **69**(3), 563–595 (2018)
48. Pishvae, M., Razmi, J., Torabi, S.A.: Robust possibilistic programming for socially responsible supply chain network design: a new approach. *Fuzzy Sets Syst.* **206**, 1–20 (2012)
49. Soyster, A.L.: Technical note—convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* **21**(5), 1154–1157 (1973)
50. Mulvey, J.M., Vanderbei, R.J., Zenios, S.A.: Robust optimization of large-scale systems. *Oper. Res.* **43**(2), 264–281 (1995)
51. Ben-Tal, A., Nemirovski, A.: Robust solutions of uncertain linear programs. *Oper. Res. Lett.* **25**(1), 1–13 (1999)
52. Ben-Tal, A., Nemirovski, A.: Robust convex optimization. *Math. Oper. Res.* **23**(4), 769–805 (1998)
53. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. *Math. Program.* **88**(3), 411–424 (2000)
54. El Ghaoui, L., Oustry, F., Lebret, H.: Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **9**(1), 33–52 (1998)
55. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. *Math. Program.* **98**(1), 49–71 (2003)
56. Yu, C.-S., Li, H.-L.: A robust optimization model for stochastic logistic problems. *Int. J. Prod. Econ.* **64**(1–3), 385–397 (2000)
57. Leung, S.C., et al.: A robust optimization model for multi-site production planning problem in an uncertain environment. *Eur. J. Oper. Res.* **181**(1), 224–238 (2007)
58. Chircop, K., Zammit-Mangion, D.: On-constraint based methods for the generation of Pareto frontiers. *J. Mech. Eng. Autom.* **3**(5), 279–289 (2013)

59. Kermani, M., et al.: A note on predicting how people interact in attributed social networks. *Int. J. Curr. Life Sci. (IJCLS)* **4**(6), 2510–2514 (2014)
60. Mesgari, I., et al.: Identifying key nodes in social networks using multi-criteria decision-making tools. In: *Mathematical technology of networks*, pp. 137–150. Springer, Berlin (2015)
61. Inderbitzen, H.M., Foster, S.L.: The teenage inventory of social skills: development, reliability, and validity. *Psychol. Assess.* **4**(4), 451 (1992)
62. Kermani, M.A.M.A., et al.: Introducing a procedure for developing a novel centrality measure (Sociability Centrality) for social networks using TOPSIS method and genetic algorithm. *Comput. Hum. Behav.* **56**, 295–305 (2016)
63. Cordeau, J.-F.: A branch-and-cut algorithm for the dial-a-ride problem. *Oper. Res.* **54**(3), 573–586 (2006)
64. Reinhardt, L.B., Pisinger, D.: A branch and cut algorithm for the container shipping network design problem. *Flex. Serv. Manuf. J.* **24**(3), 349–374 (2012)
65. Erkol, Ş., Castellano, C., Radicchi, F.: Systematic comparison between methods for the detection of influential spreaders in complex networks. *Sci. Rep.* **9**(1), 1–11 (2019)
66. Banerjee, A., et al.: The diffusion of microfinance. *Science* **341**(6144), 1236498 (2013)
67. Hu, J., et al.: A modified weighted TOPSIS to identify influential nodes in complex networks. *Physica A Stat. Mech. Appl.* **444**, 73–85 (2016)
68. Fox, W., Everton, S.: Mathematical modeling in social network analysis: using TOPSIS to find node influences in a social network. *J. Math. Syst. Sci.* **3**(10), 531–541 (2013)
69. Du, Y., et al.: A new method of identifying influential nodes in complex networks based on TOPSIS. *Physica A Stat. Mech. Appl.* **399**, 57–69 (2014)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
