ORIGINAL PAPER



Modeling the spread of infectious diseases through influence maximization

Shunyu Yao¹ · Neng Fan¹ · Jie Hu²

Received: 6 May 2021 / Accepted: 15 January 2022 / Published online: 10 February 2022 © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Mathematical approaches, such as compartmental models and agent-based models, have been utilized for modeling the spread of the infectious diseases in the computational epidemiology. However, the role of social network structure for transmission of diseases is not explicitly considered in these models. In this paper, the influence maximization problem, considering the diseases starting at some initial nodes with the potential to maximize the spreading in a social network, is adapted to model the spreading process. This approach includes the analysis of network structure and the modeling of connections among individuals with probabilities to be infected. Additionally, individual behaviors that change along the time and eventually influence the spreading process are also included. These considerations are formulated by integer optimization models. Simulation results, based on the randomly generated networks and a local community network under the COVID-19, are performed to validate the effectiveness of the proposed models, and their relationships to the classic compartmental models.

Keywords Infectious diseases spread · Influence maximization · Optimization · Integer linear programming

1 Introduction

The novel coronavirus disease (COVID-19), first identified in December 2019 in Wuhan, China, has spread rapidly to the world, resulting in the ongoing COVID-19 pandemic. In recent decades, the infectious diseases, such as severe acute respiratory syndrome (SARS), dengue fever, middle east respiratory syndrome (MERS), and Ebola virus disease, caused serious global threats. They can be easily spread from one

[☑] Jie Hu hujie_math@whu.edu.cn

¹ Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ, USA

² School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, China

individual to another, through the direct transfer of bacteria, viruses or other germs by physical touches, kisses, coughs or sneezes. Therefore, understanding the spreading or transmission process among individuals, and taking precautions to control or slow this process is critical during the pandemic periods.

Mathematical approaches have been widely utilized for modeling the spread of the infectious diseases. For example, compartmental models, including susceptible–infectious (SI), susceptible–infectious–susceptible (SIS) and susceptible–infectious–recovered (SIR) models, characterize the spread of an epidemic over time in a population of agents who pass through the states like "Susceptible", "Infectious" and "Removed" (recovered or dead). The compartmental models were first proposed by McKendrick and Kermack [1] in 1926 and were successful in predicting the behavior of outbreaks in many recorded epidemics. Recently, agent-based models (ABMs) [2,3], simulating the actions and interactions of autonomous agents, have been designed in place of simple compartmental models when trying to model precisely the phenomena occurring at the individual level. In the last two years, for COVID-19, researchers from several countries have also used mathematical modelling to predict the spread of this infectious disease [4–6] and to identify predictors of mortality [7,8].

However, these mathematical models do not explicitly consider the network structure and random transmission probability to simulate spreading process among individuals within a social network. Besides, behavior changes along the time and the cumulative effect of time that eventually influences the spreading results are seldom considered in these models.

On the other hand, to study the largest influence spread in terms of product sales or brand awareness (i.e., viral marketing) in social networks, the Influence Maximization (IM) problem is introduced as an optimization problem by Kempe et al. [9] in 2003. This problem studies a social network represented as a graph G = (V, E), where V is the set of nodes in G (i.e., individuals) and E is the set of (directed/undirected) edges in G (i.e., social links between individuals). The goal of the IM problem is to find a B-sized set of nodes (called seed set D) with the maximum influence in graph G, that is, to find the most influential individuals to maximize the influence spread $\sigma(D)$ over social networks. The influence of any seed set is defined based on some diffusion models simulating the information diffusion process like Linear Threshold (LT) model [10,11], Independent Cascade (IC) model [12], Triggering (TR) model [9], Time-Aware model [13,14], etc. The formal definition of IM is defined as follows:

Problem 1 (Influence maximization [15]) Given a graph G = (V, E) representing social network, a diffusion model on G and a budget B, find a seed set $D \subseteq V$ with $|D| \leq B$, such that the influence spread of D, $\sigma(D)$, under the given diffusion model is maximized. That is, compute $D^* \subseteq V$ such that $D^* = \arg \max_{D \subseteq V, |D| \leq B} \sigma(D)$.

In addition to viral marketing, IM is also applied in other areas, such as network monitoring [16], rumor control [17], misinformation detection [18] and social recommendation [19].

The computational hardness and algorithm results of IM under the above-defined diffusion models have also been studied widely in literature. IM problem has shown to be NP-hard under the LT, IC and TR models [9], and there exists a simple greedy algorithm that approximates the optimal solution within a factor of (1 - 1/e) for

submodular diffusion models [20]. In addition, some optimization approaches (e.g., integer linear optimization) [21–25] have also been applied to IM problem in recent years. For more details about diffusion models, hardness and algorithms of the IM problem, we refer the readers to the surveys [26–28].

There is a close similarity between the infectious disease spread and information spread in social networks. It can be seen in many corresponding concepts between IM and compartmental models, such as active/inactive nodes in IM models can be regarded as infectious/susceptible individuals in compartmental models, and the seed set of IM is considered to be initial infectious individuals in compartmental models. Besides, the diffusion process of IM is similar to the spreading process of infectious diseases over networks as well. In the diffusion framework of IM, each node $v \in V$ is associated with a status of either inactive or active. Then, based on the graph G, it considers the following diffusion process among nodes. Firstly, it starts with an initial set of active nodes (seed set $D \subseteq V$). Then, it considers the diffusion process that the seed nodes in D can "influence" their (inactive) neighbors to be active, the newly activated nodes can further activate their neighbors, and so on. This diffusion process terminates when no new nodes can be activated. Due to this similarity, the popular compartmental models of epidemiology have been adopted to study information spread in social networks as well [29-31]. Additionally, among various diffusion models of IM, Linear Threshold (LT) model is one of the most popular ones. In the LT diffusion model, influence of nodes on each other is quantized by edge weights and each node has a threshold for activation. If sum of the influence of activated (in-)neighbors of a node reaches a certain threshold, the node is activated. Recently, Cheng et al. [32] proved the targeted immunization (TI) problem [33-36], whose goal is to minimize the impact of outbreaks, is equivalent to the IM problem under the LT diffusion model and proposed an optimization framework to study outbreak minimization over networks. They gave an explicit and concise formulation of the IM problem over networks under the Time-Aware Linear Threshold model. Motivated by the above considerations, we introduce a much more general IM problem to investigate the infectious disease spreading process over a social network. The problem is defined as follows:

Problem 2 (Time-aware influence maximization problem) *Given a directed social* network $G = (V, A, \pi)$ with weight π_{ij} for each arc $(i, j) \in A$, and a budget B restricting the size of seed set, find a seed set $D \subseteq V$ such that the number of infected individuals at time T starting at D at time 0, is maximized under the threshold model.

Since individuals in the same network G may perform different behaviors during the transmission of infectious disease, to better characterize the network structure and to distinguish different behaviors, we divide the whole node set V of G into two subsets V_1 , V_2 , where individuals in different subsets take distinct actions when facing the infectious disease before any future potential behavior change, and study the interaction between V_1 and V_2 . Specifically, in the previous definition, the directed edge-weighted graph $G = (V_1 \cup V_2, A, \pi)$ represents the whole social network, where V_1 involves individuals who take precautions (e.g., wearing masks and keeping social distance) against the infectious disease while individuals in V_2 do not take active action to protect themselves from the disease. Each arc $(i, j) \in A$ indicates that there is a social link between two individuals, i.e., person *i* can directly contact person *j* through the arc (i, j). As active nodes in V_1 or V_2 may have distinct transmission probabilities to infect their neighbors, we classify this infectious disease transmission process between nodes in G into several cases. Suppose that the number of nodes is fixed all along during the transmission. The transmission probabilities π_{ij} , similar to the infection rate, are defined as

$$\pi_{ij} = \begin{cases} a, & \text{if } i \in V_1, j \in V_1, i \text{ is infectious, and } j \text{ is susceptible} \\ b, & \text{if } i \in V_1, j \in V_2, i \text{ is infectious, and } j \text{ is susceptible} \\ c, & \text{if } i \in V_2, j \in V_1, i \text{ is infectious, and } j \text{ is susceptible} \\ d, & \text{if } i \in V_2, j \in V_2, i \text{ is infectious, and } j \text{ is susceptible} \\ 0, & \text{otherwise} \end{cases}$$

where 0 < a < b < c < d < 1 in this social network *G*. Each arc $(i, j) \in A$ is associated with a positive weight π_{ij} . If there is no arc between *i* and *j* and if *i* is infectious, then $\pi_{ij} = 0$ means *i* cannot infect *j* directly. The inequality 0 < a < b < c < d < 1 represents the relationship between different transmission probabilities in various cases. For instance, a < b indicates that the transmission probability is *b* if an infectious person wears a mask and comes in contact with someone without a mask, but this probability drops to *a* if both individuals are wearing masks. Other notations are defined as follows: The (in-)neighbor set of node *i* in graph *G* is denoted by $N_G(i)$ $(N_G^-(i))$. And the degree of a node $i \in V$ in *G* is defined as $d_G(v) = |N_G(i)|$. The seed set *D* contains the initial infectious individuals.

In this paper, we will model the maximum (worst-case) infectious disease (COVID-19) propagation through influence maximization under linear threshold model. As compartmental models have become one of the most common approaches to simulate the infectious diseases, this paper follows the assumptions of compartmental models (SI, SIS and SIR) and consider three cases during infectious disease spreading:

- (1) The outbreak stage of an infectious disease (i.e., do not consider the recovery process). Transmission process is similar to the SI compartmental model.
- (2) Infections do not give immunity upon recovery from infection. That is, once a person is recovered from infection, he/she become susceptible again. Transmission process is similar to the SIS compartmental model.
- (3) Infections do give immunity upon recovery from infection. Namely, once an individual is recovered, the individual is no longer susceptible and become immune to the disease. Transmission process is similar to the SIR compartmental model.

The contribution of this paper includes: (i) We introduce the linear threshold model of IM to capture how an individual switches its status from susceptible to infectious and exploit the discrete propagation nature. (ii) We provide three explicit formulations to model the spreading of an infectious disease through IM. Behavior change and the cumulative effect of time are also considered in these models. (iii) We use the randomly generated networks and a local community network under the COVID-19 to validate the effectiveness of the proposed models. Our experiments results illustrate that the sparse and clustered structure of network topology and precautionary actions play a significant role in preventing the spread of infectious diseases.

The remainder of this paper is organized as follows. We investigate three different cases of infectious disease spreading by optimization approaches in Sects. 2, 3, and 4. In Sect. 5, we explore the behavior change in the spreading process. The experimental evaluation is illustrated in Sect. 6. Finally, Sect. 7 provides some concluding remarks.

2 Influence maximization for modeling SI spreading process

During the outbreak stage of an infectious disease, all infected individuals have not been fully recovered. The spreading process only considers the states of individuals only change from susceptible (S) to infectious (I) in only one direction, and the SI compartmental model is expressed by

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N}, \text{ and } \frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N},$$

where β is the infection rate, S(t) and I(t) represent the number of susceptible and infected individuals at time *t*, respectively, and *N* denotes the total population. In this model, each individual is considered as having the same probability of contracting the disease and contacting with same number of individuals per unit time. Each infectious individual can infect $\beta \cdot \frac{S(t)}{N}$ other susceptible ones, and thus the number of newly infected people per unit time is $\beta \frac{S(t)}{N} \cdot I(t)$.

Next, the SI spreading process is modeled within the influence maximization problem by integer optimization. For individuals changing from susceptible to infectious state, linear threshold (LT) model of diffusion process is considered.

Linear threshold (LT) is a diffusion model, first introduced by Granovetter [10], and the basic idea is that an individual can switch its status from inactive to active if a "sufficient" number of its incoming neighbors are active. In the social network G, the individual i has a threshold s_i to be infected by its infectious neighbors $j \in N_G^-(i)$, where arc (j, i) has a transmission probability π_{ji} .

Influence of nodes on each other is quantized by arc weights. If sum of the influence of activated in-neighbors of a node reaches a certain threshold, the node is activated. Specifically, given the thresholds and an initial set of active nodes, the process unfolds deterministically in discrete steps. At each time step t, a susceptible (inactive) node i becomes infectious (active) at time step t + 1 if the total weights from its active in-neighbors reach its threshold s_i , that is,

$$p_{it} = \sum_{\substack{j \in N_G^-(i) \\ j \text{ is active}}} \pi_{ji} \ge s_i.$$
(1)

where p_{it} denotes such total weights from active in-neighbors of node *i* at current time *t*. We introduce binary variables $x_{it} = 1$, $\forall i \in V$, $0 \le t \le T$, if node *i* is infectious at time *t*, and $x_{it} = 0$ otherwise, and the above transmission rules can be formulated

by the following two constraints:

$$\sum_{j \in N_G^-(i)} x_{j(t-1)} \pi_{ji} \ge s_i (x_{it} - x_{i(t-1)}), \quad \forall i \in V, t \in \{1, \dots, T\}$$
(2a)

$$\sum_{i \in N_{c}^{-}(i)} x_{j(t-1)} \pi_{ji} + \epsilon \le s_{i} + M x_{it}, \quad \forall i \in V, t \in \{1, \dots, T\}$$
(2b)

These two constraints make sure that an individual *i* is newly infected at period *t* if and only if the total influence from the infectious in-neighbors in the previous period reaches his or her threshold level. Specifically, constraints (2a) ensure that if an individual *i* is not infected at time t - 1 but become infected at period *t* (i,e., $x_{i(t-1)} = 0$, $x_{it} = 1$), then the total influence from the neighbors must exceed the threshold s_i . On the other hand, constraints (2b) guarantee that if an individual *i* is susceptible at period *t*, then the total influence from the neighbors is below the threshold s_i . Here, *M* is a sufficiently large positive number and we can choose $M = \max_i \sum_{j \in N_G^-(i)} \pi_{ji}$ in real experiments. Besides, we add a sufficiently small positive number ϵ in constraints (2b) to make sure that the node is activated when the total in-neighbor infection probability is exactly equal to the threshold. In real experiments, ϵ can be set as 0.001 to attain the precision we want. It is also noted that if the in-neighbor infection probability is never equal to the threshold, then we can set $\epsilon = 0$ in our experiments.

Sometimes, it is not enough to characterize the discrete propagation by nature just using the previous transmission rules (1). For example, if the in-degree of i in G is equal to 1, by the transmission rules (1), node i may never become activated as time passes. Hence, to avoid this case, we consider the **cumulative** effect of time in LT model so as to better describe the real propagation nature. For instance, for each node i, if we consider the total cumulative influence from the infectious neighbors of i in the last t days, then the condition of LT model becomes

$$p_{it} = p_{i(t-1)} + \sum_{\substack{j \in N_G^-(i) \\ j \text{ is active}}} \pi_{ji} \ge s_i,$$

where $p_{i0} = 0$ for all $i \in V$. To model this cumulative effect, a new parameter t_0 is introduced in the SI-LT model, meaning we consider the cumulative effect in the last t_0 days, and we use $\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji}$ to calculate the total infection probability from the infectious in-neighbors of node *i* in the last t_0 days. Then the constraints (2a) and (2b) become

$$\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji} \ge s_i (x_{it} - x_{i(t-1)}), \quad \forall i \in V, t \in \{1, \dots, T\}$$
(3a)

$$\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji} + \epsilon \le s_i + M x_{it}, \quad \forall i \in V, t \in \{1, \dots, T\}$$
(3b)

Deringer

In the beginning periods $(t < t_0)$, we set $x_{j(t-k)} = 0$ for node *j* if t - k < 0. Accordingly, the big *M* can be set as $\max_i t_0 \cdot \sum_{j \in N_G^-(i)} \pi_{ji}$ in real experiments. Finally, we formulate the IM problem by integer optimization method under the Linear Threshold (LT) model considering the cumulative effect of time:

$$[SI-LT] \quad \max \quad \sum_{i \in V} x_{iT} \tag{4a}$$

s.t. (3a), (3b)

$$\sum_{i \in V} x_{i0} \le B$$
(4b)

$$x_{i(t-1)} \le x_{it}, \quad \forall i \in V, t \in \{1, \dots, T\}$$

$$(4c)$$

$$x_{it} \in \{0, 1\}, \quad \forall i \in V, t \in \{0, \dots, T\}$$
(4d)

where the objective function (4a) is to maximize the potential number of infected people in the last period *T*. Constraints (3a) and (3b) represent LT transmission rule considering the cumulative effect of time. Constraint (4b) is budget restriction for the size of seed set *S*. Constraints (4c) ensures that if we only consider the outbreak stage of an infectious, then an infectious individual stays active as time passes. Constraints (4d) are usual binary restrictions for decision variables x_{it} .

3 Influence maximization for modeling SIS spreading process

The classic SIS compartment model considers a fixed population with only two compartments Susceptible *S* and Infected *I*, thus the flow of this model may be considered as follows:

$$S \rightleftharpoons I$$

From the above flow, we know individuals immediately become susceptible once they have recovered. Using the contact rate β from *S* to *I* and recovery rate δ from *I* to *S*, we can have the following differential equations:

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N} + \delta I(t), \text{ and } \frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \delta I(t).$$

Again, the number of newly infected people per unit time is $\beta S(t)I(t)/N$. If we assume that the probability of an infectious individual recovering in any time interval dt is simply δdt , then there are $\delta I(t)$ people recovered from the disease per unit time, hence we get the rate of change of I(t) is $\beta S(t)I(t)/N - \delta I(t)$. Furthermore, considering the population leaving the susceptible group is equal to the number of entering the infected class, we can have the first equation above.

Now we consider the SIS spreading process based on IM problem under LT. In compartmental models, we usually assume that the recovery rate δ is defined as the inverse of the duration of recovery, that is, the average recovery time for infectious

disease is $\lceil 1/\delta \rceil$, meaning that an individual is infectious for an average time period $\lceil 1/\delta \rceil$ [37]. Here $\lceil x \rceil$ is the ceiling function which is defined as the smallest integer that is not smaller than *x*. If we consider both infection and recovery process during epidemic spreading, and suppose the infections do not confer any long-lasting immunity, we need to change the constraints (3b) to

$$\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji} + \epsilon \le s_i + M(x_{it} + x_{i(t-1)}), \quad \forall i \in V, t \in \{1, \dots, T\}$$
(5)

These new constraints ensure that if the total influence from the in-neighbors of node *i* exceeds the threshold s_i , then either *i* is infectious at time *t* ($x_{it} = 1$), or *i* is recovered at time *t* ($x_{it} = 0, x_{i(t-1)} = 1$). Moreover, we also need to change constraints (4c) to

$$x_{i(t-1)} \le x_{it}, \ \forall i \in V, t \in \left\{1, \dots, \left\lceil \frac{1}{\delta} \right\rceil - 1\right\}$$

$$(6)$$

$$x_{i(t-1)} - \frac{1}{\lceil 1/\delta \rceil} \sum_{k=1}^{\lfloor 1/\delta \rceil} x_{i(t-k)} \le x_{it} \le \left\lceil \frac{1}{\delta} \right\rceil - \sum_{k=1}^{\lfloor 1/\delta \rceil} x_{i(t-k)}, \ \forall i \in V, t \in \left\{ \left\lceil \frac{1}{\delta} \right\rceil, \dots, T \right\}$$
(7)

which indicate if an individual *i* is infectious, he/she can be recovered from the disease after time period $\lceil 1/\delta \rceil$. Otherwise, an infectious individual stays active as time passes. Specifically, we consider three cases:

- At the outbreak stage ($t < \lceil 1/\delta \rceil$), an infectious individual stays active with the passage of time by Constraints (6);
- If $x_{i(t-1)}, \ldots, x_{i(t-\lceil 1/\delta \rceil)}$ all equal one when $\lceil 1/\delta \rceil \le t \le T$, then x_{it} become zero by Constraints (7);
- If not all $x_{i(t-1)}, \ldots, x_{i(t-\lceil 1/\delta \rceil)}$ equal one when $\lceil 1/\delta \rceil \le t \le T$, then an infectious individual stays active as time passes by Constraints (7).

Consequently, we obtain the following SIS-LT model:

[SIS-LT] max
$$\sum_{i \in V} x_{iT}$$

s.t. (3a), (4b), (4d)
(5)-(7)

4 Influence maximization for modeling SIR spreading process

The SIR (Susceptible–Infectious–Recovered) compartmental model is first proposed by Kermack and McKendrick [1]. This model considers a fixed population which is divided into three distinct classes: Susceptible (S), Infectious (I), and Recovered

(R). Suppose if an individual is recovered, the individual is no longer susceptible and becomes immune to the disease. The individual goes through consecutive states:

$$S \to I \to R$$

Using the same notations β and δ , Kermack and McKendrick [1] derived the following equations:

$$\frac{dS(t)}{dt} = -\frac{\beta S(t)I(t)}{N}, \quad \frac{dI(t)}{dt} = \frac{\beta S(t)I(t)}{N} - \delta I(t), \quad \frac{dR(t)}{dt} = \delta I(t),$$

where S(t), I(t) and R(t) represent the number of susceptible, infectious and recovered individuals at time *t*, separately, and *N* is the sum of these three. The first equation is the same as the first equation in SI compartmental model. For the second and third equations, consider the population leaving the infected class as equal to the number entering the recovered class. There are $\delta I(t)$ of infectious individuals leaving the infected class per unit time to enter the recovered class, which leads to the second and third equations.

Next, the SIR spreading process is modeled within the IM problem by integer optimization. New binary variables y_{it} denote whether node *i* is recovered at period *t* if $y_{it} = 1$ or not if $y_{it} = 0$. The main difference between SIS and SIR model lies on the recovery process, thus we need to update transmission rules that indicate an individual switches its status to incorporate the new state "Recovered". And then we can obtain the modified rules: if the total influence from the infectious in-neighbors of an individual exceed the threshold s_i at time *t*, then the individual is either infectious or recovered at time *t*, which can be formulate as

$$\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji} \ge s_i (x_{it} - x_{i(t-1)}), \quad \forall i \in V, t \in \{1, \dots, T\}$$
(9a)

$$x_{it} + y_{it} \le 1, \quad \forall i \in V, t \in \{0, \dots, T\}$$
(9b)

$$\sum_{k=1}^{t_0} \sum_{j \in N_G^-(i)} x_{j(t-k)} \pi_{ji} + \epsilon \le s_i + M(x_{it} + y_{it}), \quad \forall i \in V, t \in \{1, \dots, T\}$$
(9c)

Specifically, Constraints (9a)–(9b) make sure that if an individual *i* is newly infected at period *t*, then the total influence from the infectious in-neighbors in the previous period must exceed the threshold level s_i and node *i* should be unrecovered at time *t*. Since when *i* is recovered at time *t*, the total influence from the in-neighbors of *i* may still exceed the threshold s_i at this point, Constraints (9c) guarantee that if the total influence from the in-neighbors of node *i* exceed the threshold s_i at time *t*, then either *i* is infectious at time *t* ($x_{it} = 1$), or *i* is recovered at time *t* ($x_{it} = 0$, $y_{it} = 1$).

From Constraint (9b), every individual has only one of the three states (susceptible, infectious or recovered) at time *t*. The states are characterized by variables x_{it} and y_{it} , and the relationship between these variables can be ensured by (9b) and the following

constraints:

E1 (07

$$y_{it} - y_{i(t-1)} \le x_{i(t-1)}, \quad \forall i \in V, t \in \{1, \dots, T\}$$
 (10a)

$$x_{i(t-1)} - y_{it} \le x_{it}, \quad \forall i \in V, t \in \{1, \dots, T\}$$
 (10b)

Constraints (10a) ensure that if an infected individual recovers at time t, then he/she must be infected at time t - 1. Constraints (10b) indicate that how an infectious individual stays active as time passes, i.e., if a node i is infectious at previous time t - 1 and is not recovered at current time t, then node i remains infectious.

We still need some constraints to ensure the recovery process. Recall that an infectious individual becomes recovered if he/she stays infectious for a time period $\lceil 1/\delta \rceil$. Based on this fact, the whole recovery process can be guaranteed by the following constraints:

$$y_{it} = 0, \quad \forall i \in V, t \in \left\{0, \dots, \left\lceil \frac{1}{\delta} \right\rceil - 1\right\}$$
(11a)

$$\sum_{k=1}^{\lfloor 1/\delta \rfloor} x_{i(t-k)} - \left\lceil \frac{1}{\delta} \right\rceil + 1 \le y_{it}, \quad \forall i \in V, t \in \left\{ \left\lceil \frac{1}{\delta} \right\rceil, \dots, T \right\}$$
(11b)

$$\sum_{k=1}^{|1/\delta|} x_{i(t-k)} \ge \left\lceil \frac{1}{\delta} \right\rceil (y_{it} - y_{i(t-1)}), \quad \forall i \in V, t \in \left\{ \left\lceil \frac{1}{\delta} \right\rceil, \dots, T \right\}$$
(11c)

$$y_{i(t-1)} \le y_{it}, \quad \forall i \in V, t \in \{1, \dots, T\}$$
 (11d)

Constraints (11a) mean all nodes remain unrecovered at the outbreak stage (i.e., $t < \lceil 1/\delta \rceil$) of the disease. Constraints (11b)–(11c) guarantee that an infectious individual *i* recovers from the disease if and only if it takes $\lceil 1/\delta \rceil$ days. Constraints (11d) show that a recovered individual *i* remains recovered as time goes by.

To better understand the these constraints as a whole, we can divide the periods of transmission process into five cases:

- (i) A susceptible individual remains susceptible $(x_{i(t-1)} = x_{it} = 0)$: Constraints (10a) ensure that $y_{i(t-1)} = y_{it}$ and Constraints (11a) and (11c) guarantee that $y_{i(t-1)} = y_{it} = 0$. By Constraints (9c), we obtain the total influence from the active in-neighbors of node *i* does not exceed the threshold s_i .
- (ii) A susceptible individual becomes infectious $(x_{i(t-1)} = 0, x_{it} = 1)$: The total influence from the active in-neighbors of node *i* must exceed the threshold s_i by Constraints (9a). Constraints (9b) and (11d) ensure that $y_{i(t-1)} = y_{it} = 0$.
- (iii) An infectious individual remains infectious $(x_{i(t-1)} = x_{it} = 1)$: Constraints (9b) ensure that $y_{i(t-1)} = y_{it} = 0$. Constraints (9a) and (9c) become redundant. Constraints (10b) ensure that node *i* remains infectious as time passes.
- (iv) An infectious individual becomes recovered ($x_{i(t-1)} = 1, x_{it} = 0$): Constraints (9b) ensure that $y_{i(t-1)} = 0$ and Constraint (11b) guarantee that $y_{it} = 1$ as it takes $\lceil 1/\delta \rceil$ days to recover from the disease.

(v) A recovered individual remains recovered $(y_{i(t-1)} = y_{it} = 1)$: Constraints (9b) ensure that $x_{i(t-1)} = x_{it} = 0$. Constraints (11d) ensure that node *i* remains recovered as time passes.

Finally, putting the above constraints together, we obtain the following SIR-LT formulation:

[SIR-LT] max
$$\sum_{i \in V} (x_{iT} + y_{iT})$$

s.t. (4b), (4d), and $y_{it} \in \{0, 1\}$, $\forall i \in V, t \in \{0, \dots, T\}$
(9a)-(9c), (10a)-(10b), (11a) - (11d)

where the objective function $\sum_{i \in V} (x_{iT} + y_{iT})$ is to maximize the total number of infectious and recovered individuals at time *T* so as to better describe the final influence spread.

Remark 1 Note that Constraints (9b), together with Constraints (11d), enforce the assumption of SIR model, i.e., once an individual is recovered, the individual is no longer susceptible and become immune to the disease.

5 Behavior change

In Sect. 1, we introduce V_1 and V_2 to denote that individuals take distinct actions to face the infectious disease initially, that is, V_1 involves individuals who take precautions against the infectious disease while individuals in V_2 do not take active action to protect themselves from the disease in the beginning period. If some individuals in the node set V_2 may change their minds and take precautions against disease as time passes, then the above network optimization models become dynamic network models. In this section, we consider behavior change among the network G.

We still consider the graph *G* as a whole, and study the connection between nodes within *G*. To better describe each case, we introduce more notations as follows. Let G_1, G_2 be the graph induced by V_1 and V_2 , respectively. Suppose that an individual $i \in V_2$ take precautions (e.g., wears a mask) if the number of infectious neighbors of *i* exceed the threshold n_0 and we introduce new binary variables z_{it} to indicate whether node $i \in V_2$ take precautions at period *t* if $z_{it} = 1$ or not if $z_{it} = 0$. (Usually we choose $n_0 = 1$ meaning that an individual changes the behavior to take precautions if one of his/her neighbors becomes infected.) If we take SIR-LT model as an example (SI-LT and SIS-LT are similar), then constraints (9a)–(9c) should be made small modifications [see below constraints (13a)–(15e)].

An individual $i \in V_2$ change the behavior if and only if the number of infectious neighbors of node $i \in V_2$ exceed the threshold n_0 , which can be formulated by the following two constraints:

$$\sum_{j \in N_G(i)} x_{j(t-1)} \le n_0 + M z_{it}, \quad \forall i \in V_2, t \in \{1, \dots, T\}$$
(13a)

Deringer

$$\sum_{j \in N_G(i)} x_{j(t-1)} \ge n_0(z_{it} - z_{i(t-1)}), \quad \forall i \in V_2, t \in \{1, \dots, T\}$$
(13b)

Specifically, Constraints (13a) ensure that if the number of infectious neighbors of $i \in V_2$ exceed the threshold n_0 , then the individual will change the behavior. Constraints (13b) make sure that if an individual $i \in V_2$ change the mind at period t (i.e., $z_{i(t-1)} = 0$, $z_{it} = 1$), then the total number of infectious individuals must exceed the threshold n_0 .

We have some requirements for the variable z_{it} as well. Based on the definition of z_{it} and V_2 , every individual in V_2 do not take precautions initially. Further, we also assume that an individual $i \in V_2$ can only change the behavior once during the spreading process, and then we obtain the following constraints:

$$z_{i0} = 0, z_{it} \in \{0, 1\} \quad \forall i \in V_2, t \in \{1, \dots, T\}$$
(14a)

$$z_{i(t-1)} \le z_{it}, \quad \forall i \in V_2, t \in \{1, \dots, T\}$$
 (14b)

Constraints (14a) are binary restrictions for z_{it} and indicate that each node $i \in V_2$ do not take precautions at the beginning. Constraints (14b) guarantee that if a person takes precautions at some point, then he/she will always choose to take precautions as time passes.

Now we consider the SIR transmission rules incorporating the behavior change. Note that the probability π_{ji} between j and i mainly depends on (1) which set that i and j comes from (2) whether i or j take precaution at the current time period. Considering all situations, we have the following constraints:

$$\sum_{k=1}^{t_0} \left\{ \sum_{j \in N_{G_1}^-(i)} ax_{j(t-k)} + \sum_{j \in N_{G_2}^-(i)} [cx_{j(t-k)} + (a-c)x_{j(t-k)}z_{j(t-k)}] \right\}$$

$$\geq s_i(x_{it} - x_{i(t-1)}), \forall i \in V_1, t \in \{1, \dots, T\}$$
(15a)
$$\sum_{k=1}^{t_0} \left\{ \sum_{j \in N_{G_1}^-(i)} [bx_{j(t-k)} + (a-b)x_{j(t-k)}z_{i(t-k)}] + z_{i(t-k)} \right\}$$

$$\sum_{j \in N_{G_2}^-(i)} [cx_{j(t-k)} + (a-c)x_{j(t-k)}z_{j(t-k)}]$$

$$+ (1 - z_{i(t-k)}) \sum_{j \in N_{G_2}^-(i)} [dx_{j(t-k)} + (b-d)x_{j(t-k)}z_{j(t-k)}] \right\}$$

$$\geq s_i(x_{it} - x_{i(t-1)}), \forall i \in V_2, t \in \{1, \dots, T\}$$
(15b)
$$x_{it} + y_{it} \leq 1, \quad \forall i \in V, t \in \{0, \dots, T\}$$

$$\sum_{k=1}^{t_0} \left\{ \sum_{j \in N_{G_1}^-(i)} ax_{j(t-k)} + \sum_{j \in N_{G_2}^-(i)} [cx_{j(t-k)} + (a-c)x_{j(t-k)}z_{j(t-k)}] \right\} + \epsilon$$

$$\leq s_i + M(x_{it} + y_{it}), \ \forall i \in V_1, t \in \{1, \dots, T\}$$
(15d)
$$\sum_{k=1}^{t_0} \left\{ \sum_{j \in N_{G_1}^-(i)} [bx_{j(t-k)} + (a-b)x_{j(t-k)}z_{i(t-k)}] + z_{i(t-k)} \right\}$$

$$= \sum_{j \in N_{G_2}^-(i)} [cx_{j(t-k)} + (a-c)x_{j(t-k)}z_{j(t-k)}]$$

$$+ (1 - z_{i(t-k)}) \sum_{j \in N_{G_2}^-(i)} [dx_{j(t-k)} + (b-d)x_{j(t-k)}z_{j(t-k)}]$$

$$+ \epsilon \leq s_i + M(x_{it} + y_{it}), \ \forall i \in V_2, t \in \{1, \dots, T\}$$
(15e)

Constraints (15a)–(15e) represent the new dynamic transmission rules when some nodes in V_2 may change the behaviors at time t. The new transmission rules consider three cases: (a) whether the neighbor j of i is infectious at time t - k, (b) whether $j \in N_G^-(i)$ take precautions at time t - k, (c) whether i takes precautions at time t - k. Thus, we use the product terms $x_{j(t-k)}z_{i(t-k)}$ and $x_{j(t-k)}z_{j(t-k)}z_{j(t-k)}$ to identify the combination of these cases. For instance, the term $z_{i(t-k)}[cx_{j(t-k)} + (a-c)x_{j(t-k)}z_{j(t-k)}]$ in constraints (15b) means that when node $i \in V_2$ takes precautions and its neighbor $j \in V_2$ is infectious at time t - k, if j also takes precautions at time t - k, then the transmission probability from j to i is a, else the transmission probability is c. Based on the above consideration, the transmission probability π_{ji} can be determined accordingly. For instance, if $i \in V_1$, $j \in N_{G_2}^-(i)$, $x_{j(t-k)} = 1$ and $z_{j(t-k)} = 1$, then $\pi_{ji} = a$ in Constraints (15a) and (15d). Finally, the SIR-LT-dynamic model is represented as follows:

[SIR-LT-dynamic] max
$$\sum_{i \in V} (x_{iT} + y_{iT})$$

s.t. (4b), (4d), and $y_{it} \in \{0, 1\}$, $\forall i \in V, t \in \{0, ..., T\}$
(10a)-(10b), (11a)-(11d)
(13a)-(13b), (14a)-(14b), (15a)-(15e)

To linearize this model, we can introduce new binary variables $w_{ijt} = x_{it}z_{jt}$ and $h_{ijt} = x_{jt}z_{it}z_{jt}$ and consider the following linear inequalities

$$w_{ijt} \le x_{it}, w_{ijt} \le z_{jt}, x_{it} + z_{jt} - 1 \le w_{ijt}, \quad \forall i, j \in V, t \in \{1, \dots, T\}$$

$$h_{ijt} \le x_{jt}, h_{ijt} \le z_{it}, h_{ijt} \le z_{jt},$$

$$x_{it} + z_{it} + z_{jt} - 2 \le h_{ijt}, \quad \forall i, j \in V, t \in \{1, \dots, T\}$$

Deringer

to represent w_{ijt} and h_{ijt} , separately.

6 Experimental evaluation

In Sects. 2, 3, 4 and 5, the proposed models SI-LT, SIS-LT, SIR-LT and SIR-LTdynamic are all integer programming models, which will be implemented by using optimization solver CPLEX in this section. Although there was no data that can quantify how wearing a mask reduces the risk of contracting the COVID-19 by the end of 2020, according to the CDC (Centers for Disease Control and Prevention), there still exist some qualitative analyses of wearing masks to reduce the risk of disease infections [38–40]. Based on the previous research and reports, here we initialize the optimization models by choosing the parameters a = 1.5%, b = 5%, c = 30% and d = 90% to represent different risks of transmission, with thresholds $s_i = 0.99$ for each $i \in V$. Suppose one unit of time equals 1 day, we set the recovery rate $\delta = 0.04$ due to the results of the study in [41]. It is found that the average recovery time of Covid-19 patients in India is 25 days (95% CI 16–34 days). The above values of parameters were used for all experiments unless specified.

In the following, we present network transmission results for LT models on random generated Watts–Strogatz small-world graphs and a community network.

6.1 Network transmission

We use the connected Watts–Strogatz small-world graphs to simulate the propagation characteristics of the disease. For each small-world network used in this subsection, we randomly divide the graph into two networks G_1 , G_2 with equal orders (i.e., $|V_1|/|V| = 1/2$). All edges in the graphs are bidirectional, and the network simulation parameters are as follows: The number of nodes n = 50 and 100, the number of edges m = 100 and 200, the random reconnection probability p = 0.5, each node joined with its 5 nearest neighbors in an initial ring topology.

We first compute the transmission results in SI-LT model for three different budgets in two Watts–Strogatz small-world networks. Since we only consider the outbreak stage in SI-LT transmission process, we choose T = 25 for this model. Figure 1 shows the infection curves for three different budgets B = 3, 5, 9 on a connected Watts–Strogatz small-world graph of order |V| = 50, and B = 5, 10, 15 on a graph of order |V| = 100 by solving SI-LT model.

Overall, the larger the size of seed set, the faster the rate of spread, the more people infected at the same time. But there exist some exceptions like B = 10 in Fig. 1b. This situation can be attributed to the network topology and the optimization process. Since our final goal is to find the maximized influence at time T, it may not always be the optimal in any previous period t < T. Moreover, Fig. 1 also illustrates that, if we consider the epidemic spreading over a network, then the rate of spread increases faster than a linear function.

Next we compute the transmission results on the same Watts–Strogatz small-world graphs in SIS-LT and SIR-LT models. For comparison, we also compute evolution



Fig. 1 Diagram of the SI-LT model

results for the traditional SIS and SIR compartmental model (see Fig. 4). Note that when we obtain the solutions after solving SIS-LT and SIR-LT models, we also get the number of susceptible individuals at time *t* is equal to $|V| - \sum_{i \in V} x_{it}$ in SIS-LT model and $|V| - \sum_{i \in V} x_{it} - \sum_{i \in V} y_{it}$ in SIR-LT model. Figures 2 and 3 are the diagrams of infection curves of SIS-LT and SIR-LT model, separately.

From SIS-LT results in Fig. 2, we can see the curves perform similarly as SIS compartmental model at first and the value of budget *B* has little effect on the final objective function value. Further, the curves show some kinds of periodicity from about 25 days, which is very different from traditional SIS compartmental model (see below Fig. 4a). The main reason that can cause this problem is our optimization approach itself. Since we suppose that an infectious individual can be recovered from the disease after time period $\lceil 1/\delta \rceil$ in SIS-LT model, the periodicity in SIS-LT curves comes from the recovered individuals during the $\lceil 1/\delta \rceil$ time window. The declines in blue SIS-LT curves represent the recovery process of the previous infectious individuals, and this effect could become much more remarkable at about $k \cdot \lceil 1/\delta \rceil$ days (k = 1, 2, ...), since almost all individuals in the network are infectious before these time periods due to small-sized network structure and there could be less new infectious individuals at time periods $k \cdot \lceil 1/\delta \rceil$ comparing to the number of new recovered individuals.

One way to solve this periodicity problem might be increasing the order of the networks. For instance, consider a random network of 1000 nodes. In our experiments, the number of infectious individuals increases very fast at the beginning over these small network structures, which results in a phenomenon that almost all individuals become infectious at about 15 days or earlier. If the network scale can be large enough, then the number of new infectious individuals will be greater than the number of new recovered individuals at time periods $k \cdot \lceil 1/\delta \rceil$, $k = 1, 2, \ldots$ And the declining and periodical phenomenon will disappear at these time periods. However, given that optimization solver was not efficient in solving large sparse networks, methods and techniques in speeding up solving these models are still expected. Also, it's worth mentioning that this phenomenon does not appear in SIR-LT model, since once an infectious individual becomes recovered in SIR-LT model, he/she will become immune to the disease and exclude from susceptible status permanently, as shown in the green curves in Fig. 3.



Fig. 3 Diagram of the SIR-LT model

Figure 3 illustrates that although budgets are different in SIR-LT models, most curves have very similar shapes and trends. For instance, in Fig. 3b, the time when the number of infections peaked for the first time is in about 18 days, and the curves become stable in about 45 days that is much shorter than traditional SIR compartmental model in Fig. 4b.

6.2 Comparison of different edge densities

In this subsection, we use three connected Watts–Strogatz small-world graphs of different edge densities for comparison. Each node joins with 5, 7 and 10 nearest neighbors in an initial ring topology, which generates three graphs of size m = 100, 150, 250. Other parameters: the number of nodes n = 50, the random reconnection probability p = 1/2, the ratio $|V_1|/n = 1/2$, the period T = 70, the budget B = 5.

From Fig. 5, we can find out the time when the number of infections peaked for the first time becomes shorter as the edge density increases. Three curves have very similar shapes and trends whether in Fig. 5a or b. Besides, periodic oscillations are more evident in Fig. 5a than Fig. 2a.

Regarding the change of behavior on the transmission process, we compute the infection curves for the dynamic network model as well. Figure 6 shows how this



Fig. 4 Diagram of the compartmental models with initial values: total population n = 100, number of initial infectious individuals I(0) = 5, infection rate $\beta = 0.2$, recovery rate $\delta = 0.04$



Fig. 5 Comparison of different edge densities in LT optimization model with initial values |V| = 50, T = 70, $t_0 = 5$, B = 5

strategy of wearing masks affects the transmission process in SIR-LT model. Compared with SIR-LT model, the total number of infected individuals within 70 days in this dynamic SIR-LT model is much less than that of SIR-LT model. To some extent, Fig. 6 illustrates that if people can take precautions quickly to protect themselves, it will be very effective to prevent epidemic spread over the whole network.

6.3 Simulation on a community network

In this section, we simulate the infectious disease transmission process in a community located in Wuhan, China during the lockdown of this city starting from Jan 23, 2020. There are 25 buildings in this community, every buildings has 18 households and we assume that each household may have 2, 3 or 5 people and the percentages of a family made up of 2, 3 and 5 members are 20%/50%/30%. Thus there are 1512 nodes in total in this community network. Next we consider the edges between these nodes. During the lockdown of this city, people may directly contact each other when they are family, use the same elevator, or pick up groceries at the same time. Since family members



Fig. 6 Diagram of the SIR-LT dynamic network model with initial values |V| = 50, $|V_1|/|V| = 1/2$, T = 70, $t_0 = 5$, B = 5, $n_0 = 1$. **a** The susceptible, infected and recovered population fraction curves; **b** Fraction of individuals wearing masks versus time in the network *G*



Fig. 7 A community network with 1512 nodes of 25 buildings, $25 \times 18 = 450$ yellow nodes are the representatives of families, 25 green nodes are the resident representatives in each building, brown nodes represent the family members who always stay at home during the lockdown of this city

contact with each other in their homes, we consider a complete graph structure in each family. Suppose only one person in each family goes out during the lockdown of Wuhan, and each building needs one resident representative to pick up groceries outside the community. Thus, 18 people may contact when using the same elevator and 25 people may meet when they pick up groceries. We consider a cycle topology among these 18 people and 25 people separately meaning that they do these things in order.

Figure 7 shows the community network we constructed. Green, yellow and brown nodes represent resident representatives in buildings, representatives of families and family members who always stay at home, separately.

Since we only consider the outbreak stage of COVID-19 in a community, we use SI-LT model to simulate this spreading process on this network. Suppose family members who always stay at home during the lockdown of this city do not wear masks all along, others (the representatives) in this network all wear masks to protect themselves. That is, if we use $G = (V_1 \cup V_2, E)$ to denote this community network, then the node set V_1



Fig. 8 Simulation results of the SI-LT model on the community network with initial values: T = 30, $t_0 = t$, B = 3

includes all the 450 representatives in this community, the others belong to the node set V_2 . And the transmission probabilities π_{ij} can be interpreted in a more specific way. For instance, if $\pi_{ij} = a$, then it represents an infectious individual *i* can be in contact with a susceptible individual *j* and *i* and *j* are both representatives in this community. Other parameter settings are T = 30, $t_0 = t$, B = 3.

Since the network is too large to solve it directly by integer programming solver, we propose a simulation-based approach. Note that this community network we constructed has a strong symmetric structure and the size of seed set is a small value (B = 3). Three nodes in candidate seed sets may belong to some cases like three nodes are all green nodes, all yellow nodes, all brown nodes in the same family, all brown nodes in different families, two green nodes and one brown node, etc. Due to these reasons, we can first generate all possible candidate seed sets by eliminating all duplicates, then run the simulation starting at these candidate seed sets one by one and choose the best solution achieving the maximized influence spread. The results of SI-LT model can be obtained quickly, once a candidate seed set is given.

The simulation results show that the final solution consists of three brown nodes coming from three different families. The infection is clustered, and there are 15 infected people in total during the lockdown of this city, which come from three different families. And there are no infections between households when we set the parameter a = 1.5%. Figure 8 shows the simulation results on this community network, which can reveal some observations on the spreading process over community networks: (1) the infection is clustered, meaning once an individual becomes infected, his/her family can be infected as well. (2) If most individuals take precautions in a community network, then the transmission process will stop quickly. (3) The sparse and clustered structure of community network also plays an important role in preventing the further spread of infectious diseases.

6.4 Numerical results

We compare the running times and solution properties of above-mentioned optimization models and they are displayed in Table 1. The proposed models were implemented in Python 3.8 using the optimization solver CPLEX 20.1.0. All the experiments were performed on a Linux server running CentOS 7 with one AMD EPYC 7642 48-Core processor (2.3 GHz) and 512 GB memory. Computational time is reported by CPU seconds.

From Table 1, we can observe that the computation time of these models is closely related to the number of vertices in a graph and the budget we provided. Note that when n = 100, B = 5 in SI-LT, SIS-LT and SIR-LT model, the computational time is much longer than solving other instances, the reason may be that, the initial gaps of these instances (upper bound–lower bound) are larger than the other instances, and the optimization solver is not efficient in finding strong cutting plane due to the limited infection level. Furthermore, computational experiments also illustrate that methods and techniques in speeding up solving these LT models are still expected, given that the optimization solver is not efficient in solving large sparse networks. Also, when we try to consider the cumulative effect of time in the SIR-LT-dynamic model, all instances cannot be solved within an acceptable time (within 12 h), which means solving SIR-LT-dynamic model is more challenging than solving the other models and some methods of solving this model still need to be considered in the future work.

7 Conclusions

In this paper, we studied the spread of infectious disease process through the time-aware influence maximization problem. We first proposed three optimization models—SI-LT, SIS-LT and SIR-LT to investigate the discrete propagation nature and considered the cumulative effect of time in these threshold models of IM as well. Then we modeled the behavior change with respect to precautionary actions (e.g., wearing or not wearing masks) during the period of epidemic spreading over a network. In addition to studying overall infection trend like in SI/SIS/SIR compartmental models, our work also considers the interactions between individuals over networks.

Our computational experiments were performed on Watts–Strogatz small-world networks and a hand-designed community network that reflects a simplified social network in the early days of the lockdown in Wuhan, China. The results not only display several spreading curves using our approaches in networks with different settings, but also reveal several important observations on community networks. Namely, it can be found that the sparse and clustered structure of network topology and precautionary actions play a significant role in preventing the spread of infectious diseases. However, methods and techniques in speeding up solving these LT models were still expected, given that optimization solver was not efficient in solving large sparse networks. Besides, there is still a gap between the results of the SIS-LT optimization model and the actual SIS epidemic spreading process.

Future works can include the study of the impact of vaccination on the spread of the epidemic by using some optimization methods, or consideration of the interdic-

Table 1	Comparisons	of models or	1 conné	scted V	Vatts-S	trogatz small-w	orld graphs			
Model	Graphs an	d parameters				Time (s)	Optimal value	Degree sequence of S	Avg. degree of S	Avg. degree of G
	n = V	m = E	T	t_0	B					
SI-LT	50	100	25	t	3	0.74	50	[4, 5, 6]	5.00	4.00
	50	100	25	t	5	0.81	50	[4, 4, 5, 5, 7]	5.00	4.00
	50	100	25	t	6	0.58	50	[3, 4, 4, 4, 5, 5, 6, 7, 7]	5.00	4.00
	100	200	25	t	5	2823.94	66	[3, 3, 4, 6, 7]	4.60	4.00
	100	200	25	t	10	18.50	100	[2, 3, 3, 3, 3, 4, 4, 4, 6, 7]	3.90	4.00
	100	200	25	t	15	1.47	100	[2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 6]	3.80	4.00
TI-SIS	50	100	70	5	3	79.00	48	[3, 5, 5]	4.33	4.00
	50	100	70	5	5	40.92	48	[3, 4, 5, 6, 6]	4.80	4.00
	50	100	70	5	6	41.14	48	[2, 3, 3, 3, 3, 4, 5, 5, 6]	3.78	4.00
	50	150	70	5	5	84.02	49	[4, 4, 5, 5, 6]	4.80	6.00
	50	250	70	5	5	85.63	50	[6, 8, 9, 9, 10]	8.40	10.00
	100	200	70	5	5	4177.25	92	[3, 4, 4, 5, 5]	4.20	4.00
	100	200	70	5	10	542.48	92	[2, 2, 3, 4, 4, 4, 4, 4, 4, 4]	3.50	4.00
	100	200	70	5	15	192.99	92	[2, 2, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6]	4.00	4.00
SIR-LT	50	100	70	5	ю	221.41	50	[2, 3, 3]	2.67	4.00
	50	100	70	5	5	235.63	50	[2, 3, 4, 6, 7]	4.40	4.00
	50	100	70	5	6	281.27	50	[2, 3, 3, 4, 4, 5, 5, 6, 7]	4.33	4.00
	50	150	70	5	5	272.75	50	[4, 4, 5, 6, 9]	5.60	6.00
	50	250	70	5	5	732.10	50	[7, 10, 10, 11, 11]	9.80	10.00
	100	200	70	5	5	41,550.45	92	[3, 3, 4, 4, 4]	3.60	4.00
	100	200	70	5	10	3085.16	66	[2, 3, 3, 3, 4, 4, 4, 4, 4, 7]	3.80	4.00
	100	200	70	5	15	1073.18	100	[2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5]	3.47	4.00

Table 1 continued	_									
Model	Graphs ar	nd parameters	~			Time (s)	Optimal value	Degree sequence of S	Avg. degree of S	Avg. degree of G
	n = V	m = E	Т	t_0	В					
SIR-LT-dynamic	50	100	70	-	3	109.76	3	[3, 3, 4]	3.33	4.00
	50	100	70	1	5	366.94	6	[4, 5, 5, 5, 6]	5.00	4.00
	50	100	70	1	6	1197.99	12	[2,3,4,5,5,5,5,6,6]	4.56	4.00
	100	200	70	1	5	329.89	9	[3, 4, 4, 5, 7]	4.60	4.00
	100	200	70	1	10	329.94	14	[2, 3, 3, 3, 4, 4, 5, 5, 6, 6]	4.10	4.00
	100	200	70	1	15	673.47	20	[2, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 6, 6, 6]	4.07	4.00

D Springer

tion problem of epidemic spreading, that is, to minimize the maximum amount of damage that the virus could possibly inflict on the network *G*. Furthermore, there exist many other compartmental models (like SIR-Deceased (SIRD), SIR-Vaccinated (SIRV), S-Exposed-IR (SEIR), etc.) reflecting different features of disease spreading. For instance, S-Exposed-IR (SEIR) family of models encapsulate transient interactions of disease spreading and can be used for pathogen-modeling with/without fomites. These compartmental models might be considered from an optimization perspective as well.

References

- Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character 115.772, 700–721 (1927)
- Eisinger, D., Thulke, H.-H.: Spatial pattern formation facilitates eradication of infectious diseases. J. Appl. Ecol. 45(2), 415 (2008)
- Adam, D.: Special report: the simulations driving the world's response to COVID-19. Nature 580(7803), 316 (2020)
- Ndairou, F., Area, I., Nieto, J.J., Torres, D.F.: Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solitons Fractals 135, 109846 (2020)
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., Colaneri, M.: Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. 26, 855–860 (2020)
- Ivorra, B., Ferrández, M.R., Vela-Pérez, M., Ramos, A.: Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. Commun. Nonlinear Sci. Numer. Simul. 88, 105303 (2020)
- Ma, X., Ng, M., Xu, S., Xu, Z., Qiu, H., Liu, Y., Lyu, J., You, J., Zhao, P., Wang, S., et al.: Development and validation of prognosis model of mortality risk in patients with COVID-19. Epidemiol. Infect. 148, E168 (2020)
- Sousa, G., Garces, T., Cestari, V., Florencio, R., Moreira, T., Pereira, M.: Mortality and survival of COVID-19. Epidemiol. Infect. 148, E123 (2020)
- Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social net-work. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003)
- 10. Granovetter, M.: Threshold models of collective behavior. Am. J. Sociol. 83(6), 1420-1443 (1978)
- 11. Schelling, T.C.: Micromotives and Macrobehavior. WW Norton & Company, New York (2006)
- 12. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: a complex systems look at the underlying process of word-of-mouth. Mark. Lett. **12**(3), 211–223 (2001)
- Chen, W., Lu, W., Zhang, N.: Time-critical influence maximization in social networks with timedelayed diffusion process. In: Proceedings of the twenty-sixth AAAI conference on artificial intelligence. AAAI'12, pp. 592–598. AAAI Press, Toronto, Ontario, Canada (2012)
- Liu, B., Cong, G., Xu, D., Zeng, Y.: Time constrained influence maximization in social networks. In: 2012 IEEE 12th International Conference on Data Mining, pp 439–448. IEEE (2012)
- Chen, W., Lakshmanan, L.V., Castillo, C.: Information and influence propagation in social networks. Synth. Lect. Data Manag. 5(4), 1–177 (2013)
- Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 420–429 (2007)
- Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web, pp. 665–674 (2011)
- Pham, C.V., Pham, D.V., Bui, B.Q., Nguyen, A.V.: Minimum budget for misinformation detection in online social networks with provable guarantees. Optim. Lett. (2021). https://doi.org/10.1007/s11590-021-01733-0

- Ye, M., Liu, X., Lee, W.-C.: Exploring social influence for recommendation: a generative model approach. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 671–680 (2012)
- Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions-I. Math. Program. 14(1), 265–294 (1978)
- Sheldon, D., Dilkina, B., Elmachtoub, A. N., Finseth, R., Sabharwal, A., Conrad, J., Gomes, C., Shmoys, D., Allen, W., Amundsen, O., Vaughan, W.: Maximizing the spread of cascades using network design. In: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence. UAI'10, pp. 517–526. AUAI Press, Catalina Island, CA (2010). ISBN: 9780974903965
- Keskin, M.E., Güler, M.G.: Influence maximization in social networks: an integer programming approach. Turk. J. Electr. Eng. Comput. Sci. 26(6), 3383–3396 (2018)
- Güney, E.: An efficient linear programming based method for the influence maximization problem in social networks. Inf. Sci. 503, 589–605 (2019)
- Baghbani, F.G., Asadpour, M., Faili, H.: Integer linear programming for influence maximization. Iran. J. Sci. Technol. Trans. Electr. Eng. 43(3), 627–634 (2019)
- Gillen, C.P., Veremyev, A., Prokopyev, O.A., Pasiliao, E.L.: Critical arcs detection in influence networks. Networks 71(4), 412–431 (2018)
- Li, Y., Fan, J., Wang, Y., Tan, K.-L.: Influence maximization on social graphs: a survey. IEEE Trans. Knowl. Data Eng. 30(10), 1852–1872 (2018)
- Sumith, N., Annappa, B., Bhattacharya, S.: Influence maximization in large social networks: heuristics, models and parameters. Future Gener. Comput. Syst. 89, 777–790 (2018)
- Banerjee, S., Jenamani, M., Pratihar, D.K.: A survey on influence maximization in a social network. Knowl. Inf. Syst. 62, 1–39 (2020)
- More, J.S., Lingam, C.: A SI model for social media influencer maximization. Appl. Comput. Inform. 15(2), 102–108 (2019)
- Luo, W., Tay, W.P.: Identifying multiple infection sources in a network. In: 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), pp. 1483– 1489. IEEE (2012)
- Zang, W., Zhang, P., Zhou, C., Guo, L.: Locating multiple sources in social networks under the SIR model: a divide-and-conquer approach. J. Comput. Sci. 10, 278–287 (2015)
- Cheng, C.-H., Kuo, Y.-H., Zhou, Z.: Outbreak minimization vs influence maximization: an optimization framework. BMC Med. Inform. Decis. Mak. 20(1), 1–13 (2020)
- Pastor-Satorras, R., Vespignani, A.: Immunization of complex networks. Phys. Rev. E 65(3), 036104 (2002)
- Cohen, R., Havlin, S., Ben-Avraham, D.: Efficient immunization strategies for computer networks and populations. Phys. Rev. Lett. 91(24), 247901 (2003)
- Lalou, M., Kheddouci, H.: A polynomial-time algorithm for finding critical nodes in bipartite permutation graphs. Optim. Lett. 13(6), 1345–1364 (2019)
- Zhao, D., Wang, L., Li, S., Wang, Z., Wang, L., Gao, B.: Immunization of epidemics in multiplex networks. PLoS ONE 9(11), e112018 (2014)
- Keeling, M.J., Rohani, P.: Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton (2011)
- Howard, J., Huang, A., Li, Z., Tufekci, Z., Zdimal, V., van der Westhuizen, H.-M., von Delft, A., Price, A., Fridman, L., Tang, L.-H. et al.: An evidence review of face masks against COVID-19. Proc. Natl. Acad. Sci. 118(4) (2021)
- Chu, D.K., Akl, E.A., Duda, S., Solo, K., Yaacoub, S., Schünemann, H.J., El-harakeh, A., Bognanni, A., Lotfi, T., Loeb, M., et al.: Physical distancing, face masks, and eye protection to prevent person-to person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 395(10242), 1973–1987 (2020)
- Brooks, J.T., Butler, J.C.: Effectiveness of mask wearing to control community spread of SARSCoV-2. JAMA 325(10), 998–999 (2021)
- Barman, M.P., Rahman, T., Bora, K., Borgohain, C.: COVID-19 pandemic and its recovery time of patients in India: a pilot study. Diabetes Metab. Syndr. Clin. Res. Rev. 14(5), 1205–1211 (2020)

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