

RESEARCH

Open Access



Influence network design via multi-level optimization considering boundedly rational user behaviours in social media networks

Guanxiang Yun¹, Qipeng P. Zheng^{1*} , Vladimir Boginski¹ and Eduardo L. Pasiliao²

*Correspondence:

Qipeng.Zheng@ucf.edu

¹ Department

of Industrial Engineering
and Management Systems,
University of Central Florida,
Orlando, FL, USA

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

Abstract

Social media networks have been playing an increasingly more important role for both socialization and information diffusion. Political campaign can gain more supporters by attracting more mass attention and influencing them directly, while commercial campaigns can increase their companies' profits by expanding social media connection with new users. To build the optimal network structure to influence the whole, this paper studies mathematical models to simulate the users' behaviours interacting with others in the information provider's network. The behaviours of concerns include information re-posting and following/unfollowing other users. Linear threshold propagation model is used to determine the re-posting actions, Boundedly Rational User Equilibrium (BRUE) models are used to determine the following or unfollowing actions. Hence, the topology of the network changes and depends on the information provider's plan to post various kinds of information. A three-level optimization model is proposed to maximize total number of connections, the goal of the top level. The second level simulates user behaviours under BRUE. The third level maximizes the each user's utility defined in the second level. This paper solves this problem using exact algorithms for a small-scale synthetic network. For a large-scale problem, this paper uses heuristic algorithms based on large neighbourhood search. This paper also discusses possible reasons why the BRUE model may be a more accurate simulation of users' actions compared to game theory. Comparisons from the BRUE model to game theoretical model show that the BRUE model performs significantly better than game theoretical model.

Keywords: Boundedly Rational User Equilibrium, Information network, Large neighbourhood search, Linear threshold model, Budget constraint

Introduction

With the rapid development of information technology, social media platforms play a vital role in most people's life. For some commercial users or non-profit organizations, their profits can increase by extending their influence via the social media system [1, 2]. This is especially useful for important commercial users that highly depend on social media platforms. It is also true for individual users as many people are willing to expand their network connection in social media to expand their influences. The behaviour of

a user to form connection with an information provider is primarily influenced by two aspects [3]: the content of the information posted by the information provider [4], and the personality of the user who might follow the information provider. Thus, to learn how an information provider's network can be expanded, both information content and the users must be studied. Follower growth will highly depend on what kind of information providers post. We define users that determine posting plans as information providers. In this paper, we optimize the information provider's information post plan to expand its network connections. We consider an information posting plan as the decisions made during each time period regarding what content to post. Since the criteria for what is considered an interesting (and therefore, effective) post differs depending on the personality of a user, posting plans must also take user type into account.

One user can only influence another if the other user is able to view the information the original user posts. Thus, we must consider information cascading. The linear threshold model [5–7] and the independent cascading model [8] are two widely used models for information cascading. In this paper, the linear threshold propagation model is used representing information cascading. This model was first proposed by Granovetter [9] to describe people's behaviour. In this model, when the linear summation of the influence of one user's followees exceed the threshold of this user, then this user will become active. In our study, it means that when some user's followees re-post information and the linear summation of these followees' influence exceeds the threshold of this user, then this user will re-post that information.

After information cascades in the network, users then decide whether they wish to change their followers. Users determine their followees based on utility, they want to be in a network that provides them with their desired information. However, they also wish to avoid redundant information. We apply the concept of bounded rationality user equilibrium (BRUE) as the decision principle of user's for actions of connectivity. In essence, it means that users seek a connection network that achieves the high utility without exceeding the maximum utility.

This idea of BRUE originally came from Simon's Theory [10] in 1957. In this paper, human behaviour is restricted by bounded rationality. In 1972, Simon published another paper [11] which provides a fundamental illustration about the theory of bounded rationality. The theory states that human behaviour is based on a behaviour's ability to achieve a percentage of a human's optimal goals, rather than the optima, with given conditions and constraints. Simon continuously worked on bounded rationality [12, 13] to expand the application of the theory.

The concept of bounded rationality can also be applied in many fields, such as energy [14], psychology [15], military [16], transportation [17–19], etc. [20–22]. However, to the best of our knowledge, it has rarely been applied on information network systems. We propose that user actions in information networks corresponds to BRUE, that is, users are not obligated to seek the highest information utility they can get, but simply greater than a certain percentage of the maximum. Any connection plan can be chosen if it fulfills this criteria.

Some researchers used a game theoretical model in social media network to determine a user's decisions [23, 24]. However, we theorize that the BRUE model will more closely match reality than the game theoretical model. This is due to the fact that real users are

not concerned with small changes in utility. In addition, functions in network utility systems only have approximate values, thus, it is an inexact function. Based on reinforcement learning, users in the information network system will obey the BRUE principle. Finally, the BRUE model should be used over the game theoretic model since the topology of a network system changes before a user reaches a utility value equilibrium. Recent research suggests that a user's bounded rationality influences their network. Kasthurirathna and Piraveenan [25] developed a simulation for a number of strategic games. Then, they regenerated the network so that the network, on average, converged towards a Nash equilibrium, despite the bounded rationality of nodes. The link between bounded rationality distributions and social structure is important for explaining social phenomena.

We generate a three-level mathematical optimization model. The first level optimizes the information posting plan of an information provider to maximize its connections. The second level optimizes the human behaviours of other users under BRUE. It has two formats, optimistic and pessimistic conditions. In the optimistic condition, we maximize human behaviours for the connections of an information provider. In the pessimistic condition, we minimize these variables. The reason for two conditions is because, as discussed before, by introducing BRUE, users' behaviours will drop in an uncertainty set. We seek to determine how the best and the worst uncertainty will influence an information provider's network connections, which is why both conditions are necessary. The third level calculates the maximum information utility for one user, which needs to be used in second level for BRUE constraints.

We solve a small-scale synthetic network by exact algorithms. However, for large-scale networks, calculation time increases exponentially. We tackle this problem using large neighbourhood search (LNS) algorithms. It is a heuristic algorithm [26] used to solve large-scale problems. It is an effective way to find a good solution quickly when the time to find a global optimal solution is unrealistically long. This method determines the general area of a local solution and then looks within this area to find a new solution. This new solution may actually be worse than the original, however, it can get rid of the local optimal and block other possible solutions, gradually leading to better solutions.

Mathematical programming model

We propose the following model of an information network system. Our objective is to maximize our information provider's connections by controlling its information posting plan. With different posting plans, the information cascading path will be different. In addition, after information cascades, some users may connect to a new followee or disconnect from an exiting followee. Different information posting plans will lead to different numbers of followees. We utilize the linear threshold principle to determine users' information re-posting behaviour, we apply the BRUE principle to users to simulate the network reconstruction after information cascading.

In this section, we first give the introduction of the nomenclature. Second is the detailed discussion for the linear threshold principle which is used to determine whether or not a user re-posts the information. Third is the description of the mathematical model.

Nomenclature

A. Sets and indices

- N Set of users in information network system indexed by j, k, k' .
 L Set of information index by l .
 O Set of information provider index by o , this model is supposed to have only one information provider.

B. Variables

- p_{ol} Binary variable to indicate information provider o whether or not post information l .
 $p_{jl}, p_{k'l}$ Binary variable to indicate user j or user k' whether or not re-post information l .
 x_{ko} Binary variable to indicate user k whether or not to follow information provider o after information cascading.
 $x_{k'k}$ Binary variable to indicate user k' whether or not to connect to user k after information cascading.
 z_{kl} Integer variable to indicate the number of times user k receives information l .
 U_k Utility value from all information for user k .
 U_{kl} Utility value from information l for user k .
 U_k^* Maximum utility for user k can get based on the information provider's posting plan. It can be get in third-level model with fixing information posting plan.

C. Parameters

- d_{ko} Benefit for information provider o from user k if user k follows o .
 \hat{x}_{ko} Whether user k follows information provider o before information posting.
 $\hat{x}_{kk'}$ Whether user k follows user k' before information posting.

D. Functions

- $F_{kl}(\cdot)$ Function for the relation between number of times user k receives information l to the utility user k gets from information l .
 $G(\cdot)$ Function for the relation between information provider's posting plan to the user's re-posting actions. It's defined by the linear threshold principle.

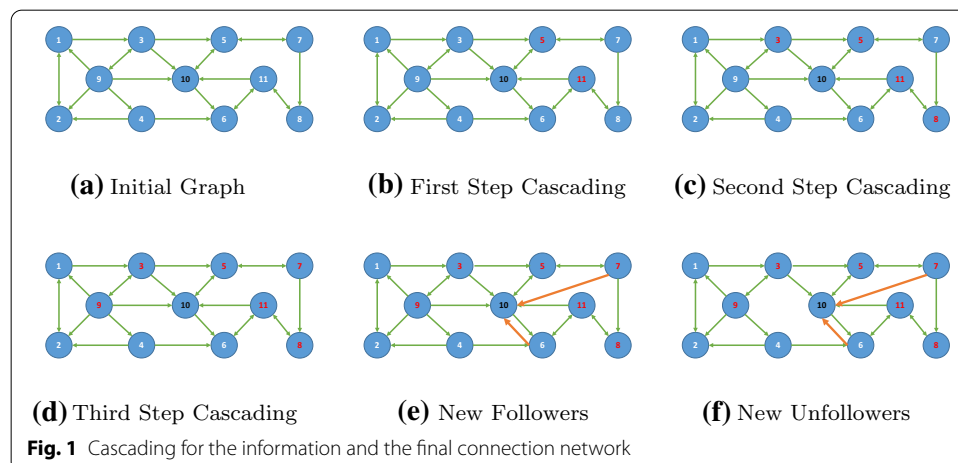
E. Concepts

- Information network The network to cascade information.
Information provider The user in the information network to post information. We suppose it has only one information provider in the network.
Users Customers in the network.
Connections The connection among users or information provider in the network.
Social network graph The topology of the information network include the users and connections.

Linear threshold model for information cascading

We use the linear threshold propagation model to determine whether a user decides to re-post information after they receive it one or several times. When the summation of the influence of one user's followees who post the information exceeds this user's threshold, it will re-post that information. The linear threshold constraint inside the mathematical CP_o model will be described later. Figure 1 gives one example of the information cascading process and the final connection network after the information cascades.

Figure 1a is the initial network. If node i has an arrow point to node j , it means node i follows node j . Some links between two users have two directions of arrows, this means that these two users follow each other. In our model, we assume it has only one information provider in the network. In this example, node 10 is our information provider. It is shown in black color. Here, we only simulate the cascading of one information. The influence and threshold values are not listed in this example. We can just suppose their values based on the following logic. This is just an example to show the cascading process based on linear threshold. Figure 1b is the first step cascading. By principle of the linear threshold, the influence of node 10 is greater than the threshold of node 5 and node 11. After node 10 posts information, node 5 and node 11 re-post it. We mark the re-post node in red color. Figure 1c is the second step of cascading. We find that node 3 and node 8 also re-post this information. Even though node 3 already follow node 10 directly in the initial network, but in the first step, the single influence of node 10 does not beat the threshold of node 3. At that time, the influence of 10 to 3 is less than the threshold of node 3. However, after the first step, node 5 also re-posts this information and node 3 receives this information from both node 5 and node 10. Currently, the summation of the influence of node 5 and node 10 is over the threshold of node 3. Thus, node 3 re-posts this information in the second step. Figure 1d is the third step of cascading. Node 9 and node 7 now re-post the information with the similar reasons of node 3. After Fig. 1d, the information cascading stops, because nodes 1, 2, 4, and 6 will no longer re-post the information. Thus, the network becomes stable. Figure 1e adds two new followers to our target node 10 after the cascading based on our BRUE model. Figure 1f shows that node 9 un-follows node 10 based on the BRUE model. We suppose the follow-unfollow action can only occur after cascading is finished for all information in



information set L . In real case, the topology may change any time, but to simplify, we just make such assumptions in our model. Figure 1e, f just gives an example after the information cascading, the topology of the network may change. The definition for users' utilities to influence the connect–disconnect action is discussed later in the mathematical programming model section.

BRUE model

BRUE is the mathematical model's equilibrium constraints from the bounded rationality. We suppose that one user in the system has multiple choices, and for choice i , it has utility $U(i)$. Without loss of generality, we can set choice i^* with the optimal utility value. Then, BRUE tells us that for any choice i has the following property that determines the possible choice for this user,

$$U(i) \geq \rho \cdot U(i^*), \quad (1)$$

where ρ is called the bounded rationality coefficient. We must have $\rho \leq 1$ because of the optimality of the choice i^* . From this constraint, we can know that by introducing BRUE to our math model, we will have an uncertainty feasible region for the users. In addition, when $\rho = 1$, it is the perfect rationality user equilibrium (PRUE), and is the Nash equilibrium. This means the user accepts the plan's utility to be the maximum utility, and thus has no motivation to move to another plan. Actually, the Nash equilibrium is a special condition of BRUE when $\rho = 1$.

In this paper, after information cascades, users may change their connections. We assume a user's choice to connect or disconnect will obey BRUE constraints, which is closer to reality when compared to the Nash equilibrium.

Pessimistic condition

The following model (CP_o) works for our information provider o to maximize the connections by determining its post plan. It is constructed under the pessimistic condition by introducing the BRUE constraints. In the pessimistic condition under BRUE constraints, the behaviour of the other users in the system acts in the worst case for our information provider. The other users' choice to connect or disconnect will lead to the minimization of the information provider's connections. The first level finds the best choice for the information provider to maximize its connections in the worst case. The second level minimizes the connections based on the users decision variables. The reason to use maximization in second level is because we want to research with the pessimistic condition to the information provider. Based on the theory of BRUE, customer's behaviour is not certainty. They should drop within their bounded rationality. Thus, in pessimistic condition, we want to find the worst influence from customers to the information provider. We want to help the provider to find a posting plan which has best customer benefit under the worst case. In second level, the decision variable is the customer's behaviour, which cannot be controlled by the information provider, but can be influenced by the information posting plan. With the certain information posting plan, each user can have the new connection or disconnection action to get their best information utilities, which is shown as U_k^* in model CP_o . But the actual action of one's decision must have the information utility that is dropped within the bounded rationality, which is defined in constraint (2h). The method to calculate the

information utility is defined from constraint (2c–2g). In this model, it also has third level, which maximizes the information utility of each user to get the U_k^* that need to be used in level 2. The details of third-level model is shown in model (TL):

(CP_o) :

$$\max_{p_{ol}} \min_{x_{ko}} BP(x_{ko}) = \sum_{k \in N} d_{ko} x_{ko} \quad (2a)$$

$$\text{s.t. } U_k = \sum_l U_{kl}, \forall k \in N, \quad (2b)$$

$$U_{kl} = F_{kl}(z_{kl}), \forall k \in N, \forall l \in L \quad (2c)$$

$$z_{kl} = \sum_{j \in N, j \neq k} x_{kj} \cdot p_{jl} + x_{ko} \cdot p_{ol}, \forall k \in N, \forall l \in L \quad (2d)$$

$$p_{jl} = G(p_{ol}), \forall j \in N, \forall l \in L \quad (2e)$$

$$x_{ko} \leq \sum_l \sum_{k' \in N, k' \neq k} p_{k'l} * x_{k'k} + \hat{x}_{ko}, \forall k \in N, \forall l \in L \quad (2f)$$

$$x_{kj}, x_{io} \in \{0, 1\}, \forall k, j \in N \quad (2g)$$

$$U_k \geq U_k^* * \rho_k, \forall k \in N \quad (2h)$$

$$U_k^* = \max_{x'_{kj}} U'_k, \forall k \in N \quad (2i)$$

$$\text{s.t. } (2b') - (2g'). \quad (2j)$$

In model CP_o , the objective function is the total connection benefit the information provider o can get from the network.

The first level's decision variable p_{ol} is a binary variable which indicates whether the information provider o will post information l to the system. This kind of variables can be controlled by the information provider.

In the second level, x_{ko} is a binary variable. When it is equal to 1, user k follows information provider o in the network. Otherwise, $x_{ko} = 0$. Constraint (2b) states that the total information utility received by user k equals the summation of the utility from all information in the system. We denote the function as $F_{kl}(\cdot)$. k is the index for customers, it is in set N . l is the index of information, it is in set L .

$$F_{kl}(z_{kl}) = \begin{cases} v_{kl} - (z_{kl} - 1) * b_{kl}, & \text{if } z_{kl} \geq 1 \\ 0, & \text{if } z_{kl} = 0 \end{cases}$$

The definition of variables and parameters in function $F_{kl}(\cdot)$ is shown in nomenclature. Figure 2a shows one example of the utility function. It means if user k receives information l 0 times. It can get 0 utility from this information. If it gets l 1 time, it can have the maximum utilities. But if it receives l multiple times. The utility from this information should decreasing linearly. The reason of decreasing is because this user should get boring by receiving the same information multiple times. This is just one assumption to the utility function. It is a logic function, we will linearize it later.

In addition, we propose another kind of utility function of all users $k \in N$ for information l . It is shown in Fig. 2b. It means when a user receives the same information multiple times, it's utility is increasing. In this case, we consider that if one user receive the information from other user's re-post, it means the other user likes this information. Then, this information can have commonality among the user's friend circle. Then, the initial value of the information will be larger if this user receive the information more times. For example, if one user has a lot of friends who re-post one information of the result of super bowl, then this user should feel this information will have more utility value to him/her. Since if he/she gets this information, it is more easy to chat about it with his/her friends. After considering this issue and the boring influence to receive same information multiple times, people with personality having strong boring feelings will obey Fig. 2a. The people with strong personality having strong feeling for friend circle will obey Fig. 2b. Actually in this case, we can see that this user has negative boring coefficient.

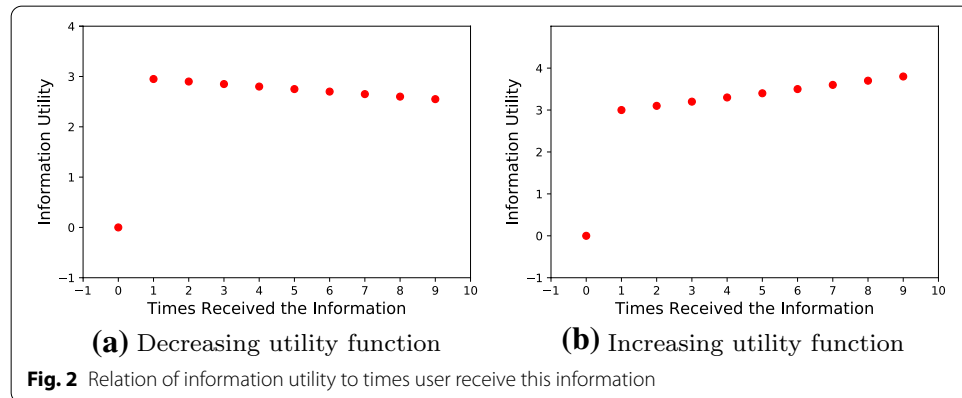
The following is the linearization of the logical function F_{kl} , this function should be applied to $\forall k \in N, \forall l \in L$:

$$\begin{aligned} \text{(LN)} \quad & F_{kl}(z_{kl}) : \\ & z_{kl} \leq M * g_{kl}, \end{aligned} \tag{3a}$$

$$g_{kl} \leq z_{kl}, \tag{3b}$$

$$U_{kl} \geq v_{kl} - (z_{kl} - 1) * b_{kl} + M(g_{kl} - 1), \tag{3c}$$

$$U_{kl} \leq v_{kl} - (z_{kl} - 1) * b_{kl} - M(g_{kl} - 1), \tag{3d}$$



$$0 \leq U_{kl} \leq M * g_{kl}, \quad (3e)$$

$$g_{kl} \in \{0, 1\}. \quad (3f)$$

These constraints give that when $z_{kl} = 0$, we will get $g_{kl} = 0$ and $U_{kl} = 0$. When $z_{kl} > 0$, we will get $g_{kl} = 1$ and $U_{kl} = v_{kl} - (z_{kl} - 1) * b_{kl}$.

Constraint (2d) shows the total times for user k to receive information l equal to the summation of user k 's followee who re-post l plus whether user k connects to o and o posts information l .

The relation of the variable p_{jl} to variable p_{ol} is restricted by the linear threshold principle, we denote their relations in function $G(\cdot)$. This function does not need to be written in formula based on our solution method large neighbourhood search will be used. The reason is once the posting plan of the information provider is decided, we can directly get the re-posting plan based on the linear threshold principle. However, the posting plan in model CP_o is the decision variable. This is the reason we write it as a constraint but do not write out the detailed constraints. We thus show the mechanism to determine p_{jl} from p_{ol} in the algorithm section. If we get the value of p_{ol} , we can directly get the value of p_{jl} by the linear threshold principle.

Constraint (2f) shows that user k does not have the choice to follow the information provider o if it does not follow user o originally and no followee of user k re-posts any information generated by information provider o . \hat{x}_{ko} is the original connection from user k to user o . If the right hand side is 0, it means user k can not get any information from o , it does not even know the existence of o , thus it can not make the choice to connect to o .

Constraint (2h) is the BRUE constraint. It states that user k can accept any follow-unfollow plan for which the information utility drops within the BRUE gaps. ρ_k is the BRUE coefficient for user k . Constraints (2i) and (2j) are the third-level model. It calculates the maximum information utility that user k can get in the system. The constraints of the third level have the same formula as the constraint (2b–2g) in the second level. However, they do not share the same variable x_{jk} and U_k . We denote all variables in level three with $*$. The third-level model is shown below in model TL :

$$(TL) : \quad U_k^* = \max_{x_{kj}'} U_k' \quad (4a)$$

$$\text{s.t.} \quad U_k' = \sum_l U_{kl}', \forall k \in N, \quad (4b)$$

$$U_{kl}' = F_{kl}(z_{kl}'), \forall k \in N, \forall l \in L \quad (4c)$$

$$z_{kl}' = \sum_{j \in N, j \neq k} x_{kj}' \cdot p_{jl} + x_{ko}' \cdot p_{ol}, \forall k \in N, \forall l \in L \quad (4d)$$

$$p_{jl} = G(p_{ol}), \forall j \in N, \forall l \in L \quad (4e)$$

$$x'_{ko} \leq \sum_l \sum_{k' \in N, k' \neq k} p_{k'l} * x'_{k'k} + \hat{x}_{ko}, \forall k \in N, \forall l \in L \quad (4f)$$

$$x'_{kj}, x'_{io} \in \{0, 1\}, \forall k, j \in N \quad (4g)$$

Model (CP2_o) gives the model to combine model (CP_o) and the linearization part.

(CP2_o) :

$$\max_{p_{ol}} \min_{x_{ko}} BP(x_{ko}) = \sum_{k \in N} d_{ko} x_{ko} \quad (5a)$$

$$\text{s.t. } U_k = \sum_l U_{kl}, \forall k \in N, \quad (5b)$$

$$z_{kl} \leq M * g_{kl}, \quad \forall k \in N, \forall l \in L \quad (5c)$$

$$g_{kl} \leq z_{kl}, \quad \forall k \in N, \forall l \in L \quad (5d)$$

$$U_{kl} \geq v_{kl} - (z_{kl} - 1) * b_{kl} + M(g_{kl} - 1), \quad \forall k \in N, \forall l \in L \quad (5e)$$

$$U_{kl} \leq v_{kl} - (z_{kl} - 1) * b_{kl} - M(g_{kl} - 1), \quad \forall k \in N, \forall l \in L \quad (5f)$$

$$0 \leq U_{kl} \leq M * g_{kl}, \quad \forall k \in N, \forall l \in L \quad (5g)$$

$$g_{kl} \in \{0, 1\}, \quad \forall k \in N, \forall l \in L \quad (5h)$$

$$z_{kl} = \sum_{j \in N, j \neq k} x_{kj} \cdot p_{jl} + x_{ko} \cdot p_{ol}, \forall k \in N, \forall l \in L \quad (5i)$$

$$p_{jl} = G(p_{ol}), \forall j \in N, \forall l \in L \quad (5j)$$

$$x_{ko} \leq \sum_l \sum_{k' \in N, k' \neq k} p_{k'l} * x'_{k'k} + \hat{x}_{ko}, \forall k \in N, \forall l \in L \quad (5k)$$

$$x_{kj}, x_{io} \in \{0, 1\}, \forall k, j \in N \quad (5l)$$

$$U_k \geq U_k^* * \rho_k, \forall k \in N \quad (5m)$$

$$U_k^* = \max_{x'_{kj}} U'_k, \forall k \in N \quad (5n)$$

$$\text{s.t. } (2b') - (2g') \quad (5o)$$

Optimistic condition

Even though we should take more concerns for the pessimistic condition, we still have a model for the optimistic condition. Under optimistic condition, the users in the system will choose the schedule which will be the best case for our information provider i . Therefore, the optimistic model ($CO2_o$) will just make change for the objective function in model ($CP2_o$).

$$\begin{aligned} (CO2_o) : \\ \max_{p_{ol}} \max_{x_{ko}} \quad BP(x_{ko}) = \sum_{k \in N} d_{ko} x_{ko} \end{aligned} \quad (6a)$$

$$\text{s.t. } (5b) - (5o), \quad (6b)$$

where all the constraints will keep the same with model ($CP2_o$).

BRUE model with budget restriction

In this model, we will consider the edit and post budget for different information. In addition, also consider in the objective function for the cost to edit and post the information. Here, just write out the model under pessimistic condition.

$$\begin{aligned} (CPB_o) : \\ \max_{p_{ol} \in \mathbf{P}} \min_{x_{ko}} \quad BP(x_{ko}) = \sum_{k \in N} b_{ko} x_{ko} - s_o \end{aligned} \quad (7a)$$

$$\text{s.t. } s_o = \sum_{l \in L} c_l \cdot p_{ol}, \quad (7b)$$

$$(5b) - (5o). \quad (7c)$$

Set \mathbf{P} is set of post plan restricted by the budget. It is defined as $\mathbf{P} = \{p_{il} | \sum_{l \in L} c_l \cdot p_{ol} \leq B_o\}$. Parameter B_o is the budget for information provider o to edit and post information. Parameter c_l is the cost to post information l . Parameter s_o is the total information post cost for o .

Algorithm

We use the linear threshold principle to determine the information cascading process. When the summation of the influence of user j 's followees who post information l is greater than the threshold of user j for information l , user j will re-post information l .

The following algorithm is to determine user's re-post decision based on the linear threshold principle and known post plan of information provider o .

Algorithm 1 ALG – LT

```

1: for  $j \in N$  do                                ▷ Start Customer Set Loop, Set all re-posting variable  $p_{jl}$  to 0
2:   for  $l \in L$  do                                ▷ Start Information Set Loop
3:      $p_{jl} \leftarrow 0$ 
4:   end for                                        ▷ End Information Set Loop
5: end for                                          ▷ End Customer Set Loop
6: for  $l \in L$  do                                ▷ Start Information Set Loop
7:   if  $p_{ol} = 1$  then
8:     for  $j \in N$  do                                ▷ Start First Customer Set Loop
9:       for  $k \in N$  do                                ▷ Start Second Customer Set Loop
10:        if  $\sum_{s \in N} \hat{x}_{ks} \cdot p_{sl} \geq T_{kl}$  and  $p_{kl} = 0$  then
11:           $p_{kl} = 1$                                 ▷ Set re-post variable to 1 is threshold is activated
12:        end if
13:      end for                                    ▷ End Second Customer Set Loop
14:    end for                                    ▷ End First Customer Set Loop
15:  end if
16: end for                                          ▷ End Information Set Loop

```

Remark 1

By applying the process of cascading with threshold principle n times, the cascading must stop.

Prove: 1, we can not just apply the cascading process one time. The reason is for example, user k cannot be activated with the first time cascading, but one of its followee gets activated. Then, perhaps with the summation of this followee's influence and the primary influence, user k can be activated. Thus, we need to apply multiple times for the cascading until the influence for re-post in the network becomes stable.

2, If after one round cascading, the information post decision of all users keep the same. Then, it means the cascading already finished, there will have no extra cascading. Even if we apply more rounds of cascading under this condition, the re-posting behaviour among users will not change any more. If after one round cascading, some information post decisions of all users have changes. It means at least one user changes their decision, so if we run the cascading process for n times. In addition, each time, it has at least one change in re-posting behaviour. The re-post behaviour must be changed more than n times. But we know the behaviour can change at most n times since it just n users in the network. Thus, either the re-posting already becomes stable before running n times or every users choose to re-post the information. No matter under which condition, the

re-posting must get stable after n times to run the cascading. Number n is the number of total users in the system.

Due to Remark 1, we run the cascading n times with the for loop at line 9 in algorithm 1.

In model $(CP2_o)$, it is a three-level optimization problem, and in constraint (5i), it has the quadratic terms. However, we can notice that if we know the value of variables p_{ol} in the first level, the problem will decompose to several one-level mixed linear integer program (MILP) problem. We can imagine if we have the value of $\{p_{ol} | \forall l \in L\}$, we can get all values of $\{p_{jl} | \forall j \in N, l \in L\}$. Then, constraint (5i) becomes a linear constraint. In addition, we can also calculate the value for the third-level problem. It means we can get the value of $\{U_k^* | \forall k \in N\}$. Then, the total problem is decomposed to (n) one-level MILP problem in the third level. In addition, (1) MILP problem in the second level, where n is the number of total users in the system.

One possible method to solve this problem is to numerate all possible plans for the first-level problem. However, we can know that the number of different schedules to post information for user i is $2^{|L|}$, where $|L|$ is the number of information our information provider may post. When $|L|$ is increasing, the number of schedule will increasing exponentially. It is not a good method for the problem with large number of information to decide whether or not to post.

For large-scale problem, we use the idea of large neighbourhood search method for the first level. The detailed algorithm is shown in algorithm 2 ALG-LNS. These two methods can be used for both pessimistic and optimistic conditions. The following algorithm just

Algorithm 2 ALG – LNS

- 1: Set $k = 0$, $\{p_{ol}^k = 0, | \forall l \in L\}, \{h[s] = 0 | \forall s \in S\}$
 - 2: Let $\{p_{ol} = p_{ol}^k | \forall l \in N\}$, calculate constraint (5n,5o) in model $(CP2_o)$ to get $\{U_k^* | \forall k \in N\}$, implement this value to BRUE constraint (5m) to get objective function value η^k . Set $h[p_{ol}^k] = 1$.
 - 3: Set $m = 0$.
 - 4: Let $\{p_{om}^{k,m} = 1 - p_{om}^k\}$, $\{p_{on}^{k,m} = p_{on}^k | \forall n \in L, n \neq m\}$. Let $\{p_{ol} = p_{ol}^{k,m} | \forall l \in N\}$. Use it to calculate constraint (5n,5o), use the solution for BRUE constraint (5m) to get the objective function value $\eta^{k,m}$.
 - 5: If $m < |L| - 1$. $m = m + 1$, go to step 4. Otherwise, go to step 6.
 - 6: $\eta^{k+1} = \{\max_m \eta^{k,m} | s.t. h[p_{ol}^{k,m}] = 0\}$, $m^* = \{\argmax \eta^{k,m} | s.t. h[p_{ol}^{k,m}] = 0\}$. Let $\{p_{on}^{k+1} = p_{on}^{k,m^*} | \forall n \in N\}$
 - 7: $n = \sum_{m \in N} h[p_{ol}^{k,m}]$.
 - 8: If $n < |L|$ and $k \leq K$ and $ct \leq TL$, $k = k + 1$, go to step 2. Otherwise, stop.
-

shows the pessimistic condition.

In algorithm 2 ALG-LNS, S is the set for all possible schedule to post information of user i . The number of elements in S is $2^{|L|}$. K is the number of iterations needed for the problem. we can know that the calculation complexity for it is $O(|L| * K)$. Based on different requirements of accuracy, we can use different K . The value $h[s]$ is an indicator whether the plan s already seems as the location maximum before. ct is the calculation time. TL is the time allowance for calculation.

From the algorithm, we can know that within each iteration k , the post plan p_{ol} is actually becoming an input. In addition, based on the linear threshold principle, we can get all of the variables' value p_{il} in the math model. Therefore, all the users' post behaviour is input now. This is the reason we do not need to linearize constraint (5i).

Computation result

Dataset

We use the synthetic network to calculate the model. We generate the network with different types of information, different numbers of users in the network. The following Table 1 gives the structure of the network. We have 3 different types of dataset to be calculated.

The coefficient b_{kl} is the coefficient for user k to get boring about information l . We generate it by using the random function in $[0, 2]$. The coefficient v_{kl} is the coefficient for user k to get how much value of the information l if just receive it for one time. We generate it randomly in $[0, 10]$. The coefficient d_{ko} is the importance of user k to provider o . We generate it using the random function in $[0, 100]$. The threshold values for each user to re-post the information are generated in two categories. First category includes the user with high threshold. With probability of 40% in total users. we generate them randomly in $[2, HT]$. The second category includes the users with low threshold. We generate them randomly in $[2, LT]$, where $HT = 10000 * N * D$ and $LT = N * D$. N is the number of users in the network. D is the density of arcs in the network. The influence between two users is generated randomly between $[1, HI]$, where $HI = N * D/2$. The posting cost for information l is c_l . It is generated randomly between $[0, 300]$. The total budget for the posting cost is \$2000.

All models and algorithms are programmed in Python 3.7 and solver Gurobi 7.5.2. All results are calculated on a personal computer. It has Intel Core i7-7820 HQ processors with CPU at 2.90 GHz and RAM with 24.0 GB.

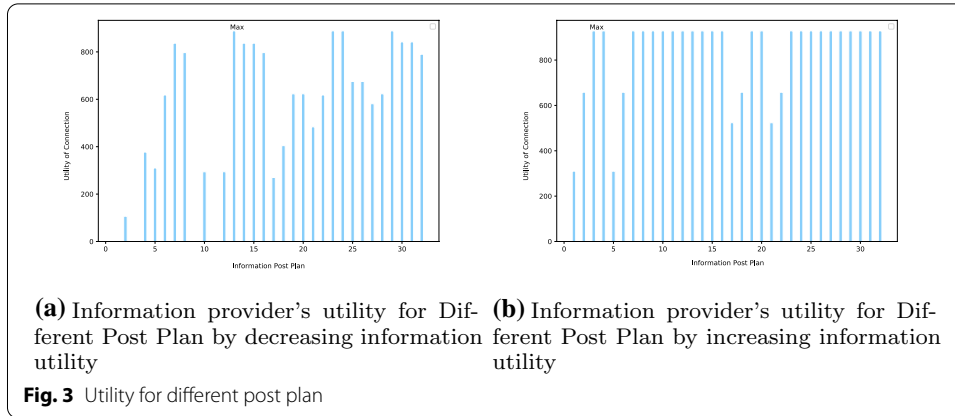
Result of connection utility for information provider

Post plan of information provider

Figure 3a gives out the maximum utilities for different post plan under BRUE coefficient $\rho = 0.8$ using the dataset under case 1 shown in Table 1. It uses the information utility

Table 1 Dataset

Case	Node	Information	Density
1	20	5	0.3
2	100	20	0.1
3	1000	5	0.1

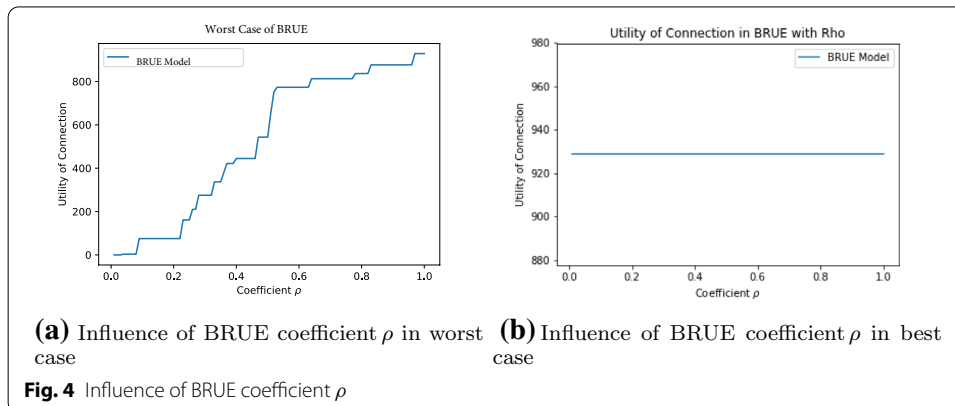


function in Fig. 2a. We can find out that maximum plan has the index 13, which means the post plan is $p_{max} = [0, 1, 1, 0, 0]$. Under this case, the maximum utility for post plan is to post information 2 and 3 and not post information 1, 4, and 5 for our information provider. This is the result for a simple example that can be enumerate all schedules to remove the first level in our model. But in large-scale problem, we will use algorithm 2 ALG-LNS to give out the heuristic solutions.

Figure 3b gives out the result using the information utility function in Fig. 2b. We can see that under this case, more post plans can get the maximum utility of connection for the information provider. It means the users in the system will easily choose connecting to others because this user can have more information utility when he or she will get the information more times.

Influence of BRUE coefficient ρ

Figure 4 shows the result of the influence of BRUE coefficient ρ in worst condition and best condition with case 1 in Table 1. We can see that under worst condition, when ρ is increasing, the total utility is also increasing. This is because by increasing of ρ , the feasible region of BRUE constraint is decreasing. For worst case, the objective function value is increasing. In addition, for best case is a horizon line, this is because when $\rho = 1$, it already get the maximum utility under this condition. In addition, the values of utility under best case and worst are also equal when $\rho = 1$. Which means under the condition game theory, worst case and best case are the same case.



Compare BRUE model and game theoretical model

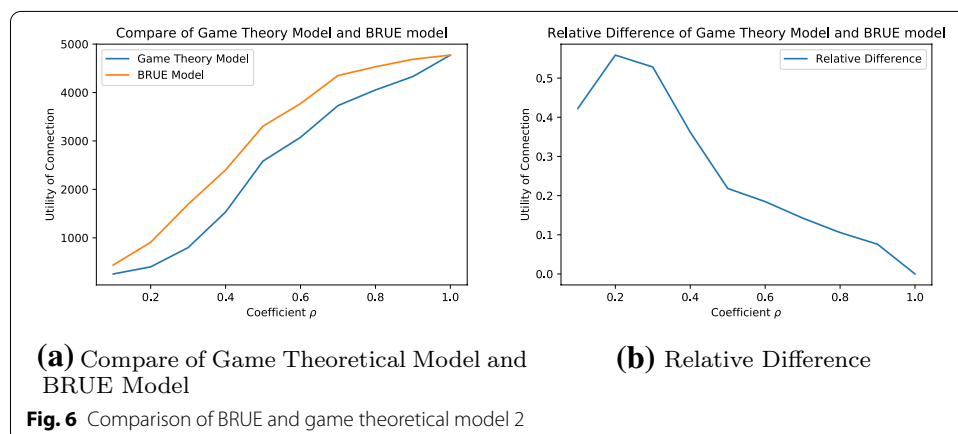
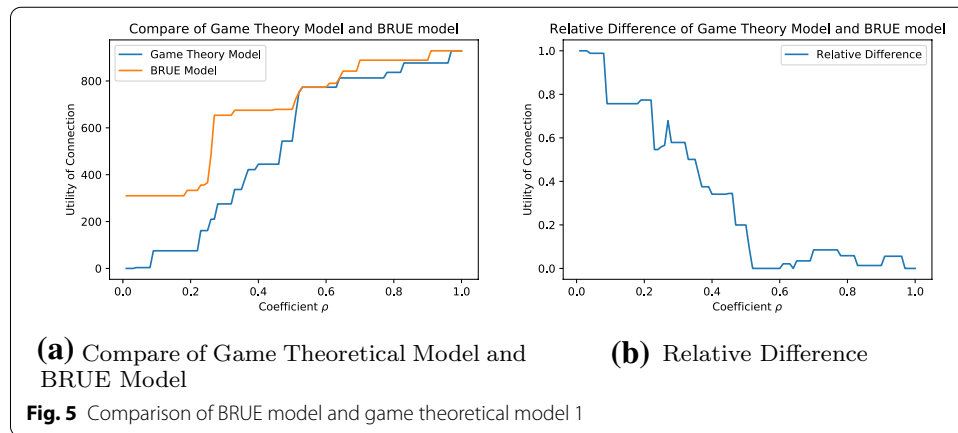
In this section, we calculate the post plan for our information provider if we use the game theoretical model to forecast the users' behaviour in the network under the worst condition. Then, we use this post schedule in our BRUE model, under different BRUE coefficients, we will get different maximum or minimum connection utilities for our information provider. We compare the relative difference to show how much the system will perform better using BRUE instead of game theoretical model under different BRUE coefficients.

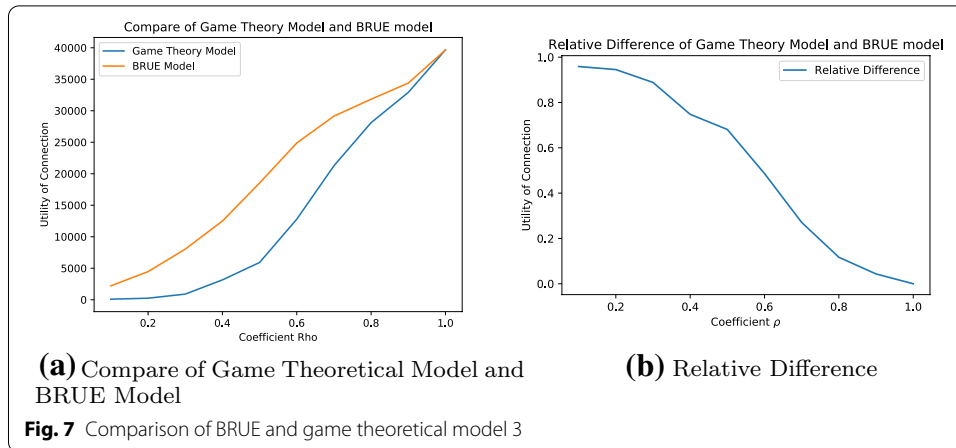
Figure 5 is the comparison the connection utility of BRUE and game theoretical model with dataset case 1 in Table 1 under different BRUE coefficient ρ . We can see that using the post behaviour got from game theoretical model and use it in BRUE, the total utility is less or equal to the post behaviour directly obtained from BRUE model. By decreasing of ρ , the relative difference generally becomes larger.

Figures 6, 7 are the solutions that is obtained relatively from case 2 and case 3 in Table 1. We can see that the larger the number of nodes in the system, the more smooth the curves are.

Result using large neighbourhood search method

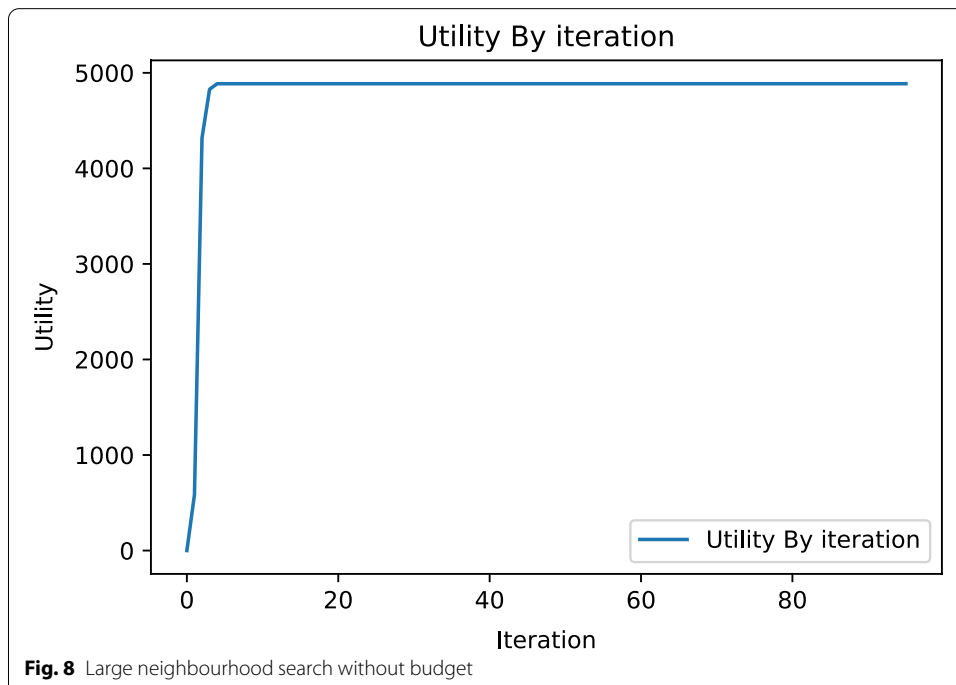
We calculate the result using the dataset in case 2. In addition, we also use different initial solutions when use large neighbourhood search.





Result without budget and cost penalty

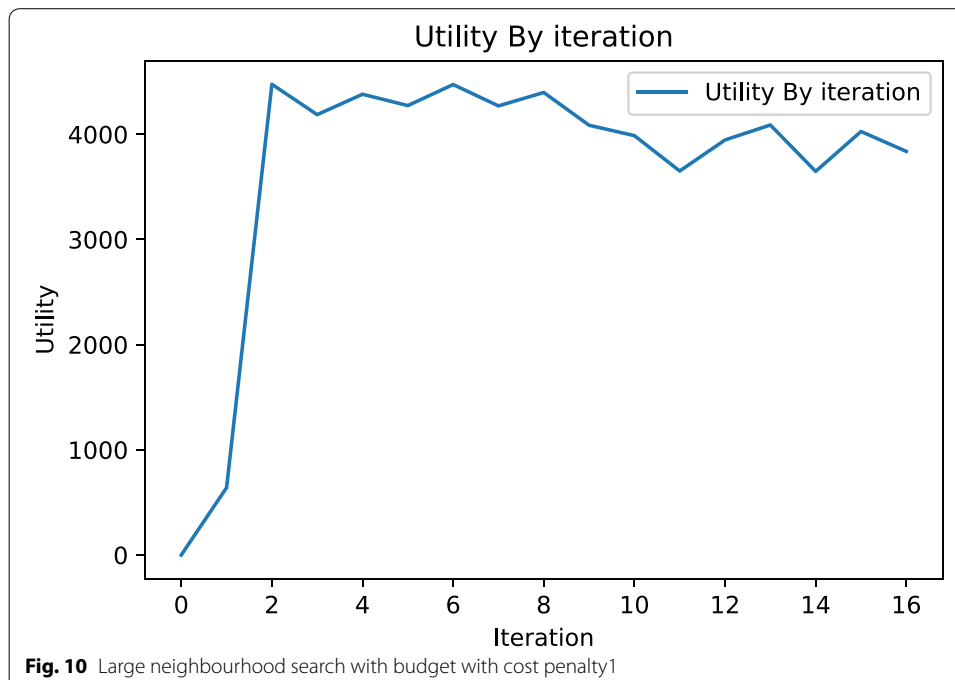
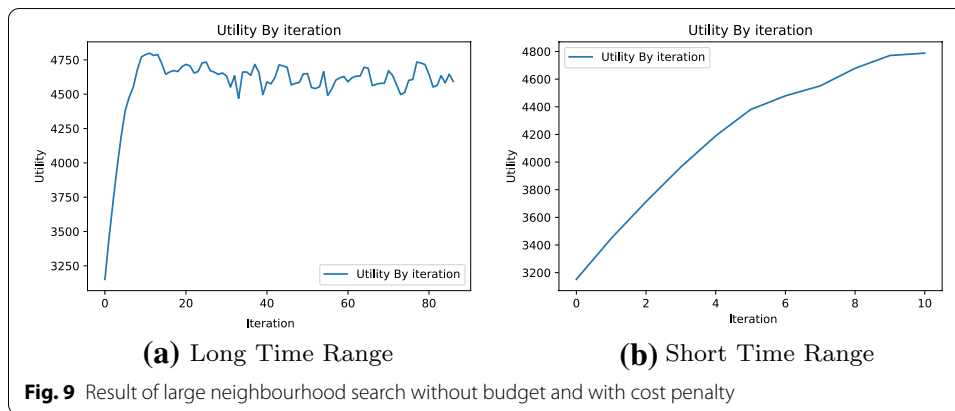
Figure 8 is the result without budget and penalty in the objective function using large neighbourhood search method under case 2. The value is the same after about iteration 8 is the same is because for different information post schedule, it may lead to the same connection of the network. Then, the result of different iterations in large neighbourhood search just keep the same. This result is obtained from the initial solution 0, which means the information provider will not post any information at the first iteration. The calculation time we set is 1 h. We can also get the result from enumerate method, which is to enumerate all possible post schedules. Compared to the large neighbourhood search method, it gets the same solution. But the calculation time is about 2 days. It is much longer than LNS.



Result with budget and cost penalty

Figure 9a is the result from data case 2. We solve it from the starting point '1', which means the information provider posts all information at the first iteration. We can see the optimal solution actually already obtained at iteration 10. In addition, the optimal solution value is 4798.84. The time cost for the first one is 1 h. For the second one, it is just 10 min that can get the optimal solution. After we enumerate all possible schedules, we also get the global optimal solution is 4798.84. But the time cost is about 2 days, which is much longer than the time using large neighbourhood search method.

Figure 10 is the result with budget and cost penalty, and the starting point is '0'. It means the first post plan is to post nothing to the network. We can first know that the total objective value is smaller than the model without budget. It is obviously because in this case actually some post plan is prohibited by the budget.



The starting point from '0' can get the optimal solution more quickly than start from '1'. This is because under this example, the final post schedule does not need to post too much information to get the optimal solution.

Conclusion

This paper develops several multi-level optimization models, based on boundedly rational user behaviours, to calculate the best posting plan for information providers seeking to build the strongest influence network. Using the best plans, the information provided can maximally expand its connections with users. Through extensive numerical experiments, we found that in the BRUE model, the smaller BRUE coefficient ρ , the less connections our information provider has in pessimistic conditions. Thus, no matter what kind of information the information provider posts, it is easier to lose such users in the network. Therefore, the information provider should pay more attention to the users with high BRUE coefficient ρ . The BRUE model performs better than the game theoretical model for maximizing the information provider's connections, especially when BRUE coefficient ρ is relatively small. In addition, we found that the large neighbourhood search method is a useful algorithm for large-scale problem, consuming a reasonable computational time. The local optimal from LNS is equal to the global optimal solution in our case study, but much less time is spent in obtaining the optimal solution. Another advantage using LNS is seen when the variable of the first level is fixed. The three-level optimization problem can be directly decomposed to several single-level MILP problems. The initial solution when using LNS can also impact convergence speed to an acceptable local optimal solution. In general, when we have more people with a high boring coefficient, it is better to start from the plan that posts zero information. But when we have more people with a low boring coefficient or negative coefficient, it is better to start from the posting plan that posts all information within the cost budget.

Acknowledgements

This material is based upon work supported by the AFRL Mathematical Modeling and Optimization Institute. In addition, we would like to thank Luke Fuller for assistance with detailed proof reading and comments that greatly improved the manuscript.

Further information

A preliminary version of this journal paper appeared as a proceeding article in: Tagarelli A., Tong H. (eds) Computational Data and Social Networks. CSoNet 2019. Lecture Notes in Computer Science, Vol. 11917. The current journal paper has since improved substantially.

Authors' contributions

GY conducted the experiments. All authors read and approved the final manuscript.

Funding

The work was supported in part by the U.S. Air Force Research Laboratory (AFRL) award FA8651-16-2-0009.

Availability of data and materials

This paper uses synthetic data. N/A.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, FL, USA. ² Air Force Research Laboratory, Eglin AFB, FL, USA.

Received: 7 April 2020 Accepted: 16 December 2020

Published online: 08 February 2021

References

1. Lovejoy K, Saxton GD. Information, community, and action: how nonprofit organizations use social media. *J Comput Med Communica*. 2012;17(3):337–53.
2. Paniagua J, Sapena J. Business performance and social media: love or hate? *Busin Horizons*. 2014;57(6):719–28.
3. Van Dijck J. The culture of connectivity: a critical history of social media. Oxford: Oxford University Press; 2013a.
4. Van Dijck J. Facebook and the engineering of connectivity: a multi-layered approach to social media platforms. *Convergence*. 2013;19(2):141–55.
5. Goyal A, Lu W, Lakshmanan LV. Simpath: an efficient algorithm for influence maximization under the linear threshold model. In: *Data Mining (ICDM), 2011 IEEE 11th International Conference On*. IEEE; 2011. p. 211–20.
6. Chen W, Yuan Y, Zhang L. Scalable influence maximization in social networks under the linear threshold model. In: *2010 IEEE International Conference on Data Mining*. IEEE; 2010. p. 88–97.
7. Borodin A, Filmus Y, Oren J. Threshold models for competitive influence in social networks. In: *International Workshop on Internet and Network Economics*. Springer; 2010. p. 539–50.
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. p. 137–146.
9. Granovetter M. Threshold models of collective behavior. *Am J Soc*. 1978;83(6):1420–43.
10. Simon HA. Models of man; social and rational. 1957.
11. Simon HA. Theories of bounded rationality. *Decis Organizat*. 1972;1(1):161–76.
12. Simon HA. Bounded rationality and organizational learning. *Organizat Sci*. 1991;2(1):125–34.
13. Simon HA. Models of Bounded Rationality: Empirically Grounded Economic Reason, vol. 3. Cambridge: MIT press; 1982.
14. Wüstenhagen R, Menichetti E. Strategic choices for renewable energy investment: Conceptual framework and opportunities for further research. *Energy Policy*. 2012;40:1–10.
15. Kahneman D. Maps of bounded rationality: psychology for behavioral economics. *Am Econ Rev*. 2003;93(5):1449–755.
16. Roberts KH, Stout SK, Halpern JJ. Decision dynamics in two high reliability military organizations. *Managem Sci*. 1994;40(5):614–24.
17. Mahmassani HS, Chang G-L. On boundedly rational user equilibrium in transportation systems. *Transportat Sci*. 1987;21(2):89–99.
18. Di X, Liu HX, Pang J-S, Ban XJ. Boundedly rational user equilibria (brue): mathematical formulation and solution sets. *Transportat Res Part B Methodol*. 2013;57:300–13.
19. Lou Y, Yin Y, Lawphongpanich S. Robust congestion pricing under boundedly rational user equilibrium. *Transportat Res Part B Methodol*. 2010;44(1):15–28.
20. Rötheli TF. Boundedly rational banks contribution to the credit cycle. *J Soc Econ*. 2012;41(5):730–7.
21. Han K, Szeto W, Friesz TL. Formulation, existence, and computation of boundedly rational dynamic user equilibrium with fixed or endogenous user tolerance. *Transportat Res Part B Methodol*. 2015;79:16–49.
22. Elsadany A. Competition analysis of a triopoly game with bounded rationality. *Chaos Solitons Fractals*. 2012;45(11):1343–8.
23. Singh VK, Jain R, Kankanhalli MS. Motivating contributors in social media networks. In: *Proceedings of the First SIGMM Workshop on Social Media*. ACM; 2009. p. 11–18.
24. Su Z, Xu Q, Fei M, Dong M. Game theoretic resource allocation in media cloud with mobile social users. *IEEE Transact Multimedia*. 2016;18(8):1650–60.
25. Kasthurirathna D, Piraveenan M. Emergence of scale-free characteristics in socio-ecological systems with bounded rationality. *Sci Rep*. 2015;5:10448.
26. Laporte G, Musmanno R, Vucaturio F. An adaptive large neighbourhood search heuristic for the capacitated arc-routing problem with stochastic demands. *Transportat Sci*. 2010;44(1):125–35.

Publisher's Note

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