Information Sciences 572 (2021) 1–15

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Misinformation influence minimization problem based on group disbanded in social networks ☆

Jianming Zhu [a,*], Peikun Ni [a], Guoqing Wang [a], Yuan Li [b]

[a] School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China
[b] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Article history:
Received 23 January 2021
Received in revised form 19 April 2021
Accepted 27 April 2021
Available online 2 May 2021

Keywords:
Misinformation
Group disbanded strategy
Echo Chamber Effect
Social networks

ABSTRACT

The booming development of online social media has changed the way people post and access information. The authenticity of content is weakened, and all kinds of misinformation on social media spread rapidly. In Online Social Networks (OSN), users arbitrarily form private groups/communities, which greatly increase the exposure rate of misinformation. Considering that echo chamber effect of groups wildly exists, this paper studies the disbanding strategy of private groups in OSNs to Minimize the Spread of Misinformation under the effect of Echo chamber effect (MSME). Given a directed acyclic OSN $G(V, E, C)$, C denotes a set of private groups, the problem of MSME is to select K groups from C, such that the spread of misinformation will be minimized by disbanding these groups. We prove the problem of MSME is NP-hard, then prove that the objective function computation of the problem of MSME is #P-hard. It is proved that the objective function of the problem of MSME is neither a submodular nor a supermodular. A greedy algorithm is constructed and several heuristic algorithms are proposed to solve the objective function which is non-submodular and non-supermodular. Our experimental simulation on four real world datasets verifies the effectiveness of our constructed algorithm.

1. Introduction

The rapid development of mobile instant messaging devices has changed the way information is disseminated [1]. People broadly participate in online social media and can freely get and post information, and truly become the main source of information interaction. People express their intrigued or views on something by releasing or posting unconfirmed content on online social medias, which provides more effective and low cost support for the dissemination of negative contents (Misinformation, Cyber violence, Heresy, etc.). Nowadays, some eye-catching misinformation on Online Social Networks (OSNs) is enough to evolve into online public opinion, affect people's daily life, and even trigger social unrest. For instance, in October 2017, when a devastating bushfire broke out in California, the government not only needed to rescue the inhabitants, but also needed to deal with rumors about timberland fires in OSNs. After the outbreak of the novel coronavirus pneumonia epidemic (COVID-19) in Wuhan, China in early January 2020, the masses in some areas of China panicked. However, the panic was not mainly caused by the coronavirus, but by a large number of fake news on various social platforms. The wide spread

☆ This work was supported in part by National Natural Science Foundation of China (NSFC) under Grant Nos. 72074203.
* Corresponding author.
E-mail addresses: jmzhu@ucas.ac.cn (J. Zhu), nipeikun18@mails.ucas.ac.cn (P. Ni), wgq94@ucas.ac.cn (G. Wang).

https://doi.org/10.1016/j.ins.2021.04.086
0020-0255/

of misinformation on OSNs has seriously affected the normal social order and has also caused interference with the correct implementation of the Chinese government's plan to fight the epidemic, etc.

In OSNs, users may not only receive misinformation from friends who have direct contact, but also may get misinformation from the groups they join. In other words, users may be influenced by friends with whom they have previously established contacts, but also be influenced by group members who they have no direct contact with, but one part of private groups with them. Users arbitrarily form private groups/communities on OSNs, which greatly increase the exposure rate of misinformation, improve the frequency of misinformation interaction between users, and expand the spreading range of misinformation. Considering this aspect, it is necessary to propose an effective strategy to diminish the overall influence of the group, thereby reducing the exposure rate of misinformation and the speed of dissemination of misinformation. However, the current misinformation control strategies only consider the influence of users by their friends, and ignore the indirect influence of other users in their group. Based on this, this paper considers the overall effect of groups on its users, and studies the disbanding strategy of private groups/communities in OSNs to minimize the spread of misinformation.

1.1. Related work

Although there has been a large amount of work on the influence maximization problem of OSNs, there is less research on the controlling of the spread of misinformation. In this section, we briefly review the relevant research work on misinformation control strategies in OSNs. Taking into consideration the topological structure of social networks [2], the utility experience of users [3] and the user attributes [4,5], many scholars [6,7] put forward the active control strategies of blocking nodes or links to minimize the influence of the misinformation. Wang et al. [3] comprehensively considered topic dynamics, Ising model, ability model and survival theory, etc. and proposed a dynamic rumor influence minimization model with user experience. Greedy algorithm and dynamic blocking algorithm are developed to minimize the number of users who accept rumors. Zhang et al. [6] used blocking measures to prevent the generation of rumors, thereby minimizing the influence of rumors on OSNs within the user's tolerance. Zhu et al. [7] introduced a new activity minimization of the misinformation influence problem, which is to select nodes from a given social network to block, to minimize the total amount of misinformation interaction between nodes. Kimura et al. [8] solved the problem of minimizing the spread of unwelcome things (such as computer viruses or malicious rumors) by blocking a limited number of links in OSNs, and proposed a good approximate solution to the problem based on natural greedy strategy.

Considering that the strategy of blocking nodes or links to control the dissemination of misinformation may have a negative impact upon the user experience in OSNs, some researchers [9–11] have proposed the strategy of publishing positive information or strategies of reflexive rumors to limit the spread of misinformation. Hosni et al. [12] considered the individual and social behaviors in OSNs, and proposed an individual behavioral statement that simulates damped harmonic motion for rumors, constructed a new rumor propagation model, aiming to minimize the influence of rumor in multiple social networks from the perspective of network reasoning and survival theory. Arazkhani et al. [10] built a multi-campaign independent cascade model based on the concepts of passive and active competition in OSNs, and proposed a centrality measure-based IBM algorithm to find a suitable subset of candidate nodes for positive diffusion, to minimize the adverse effects of misinformation. Yang et al. [11] proposed a linear threshold model with one direction state transition to simulate two different types of competitive information propagation in the same network, and proposed a heuristic method based on diffusion dynamics to solve this problem. Furthermore, a few scholars comprehensively considered the advantages and disadvantages of anti-rumor campaign strategies and blocking nodes/links strategies to study misinformation control strategies. Hosni et al. [13] considered the advantages and disadvantages of the anti-rumor campaign strategy and the blocking nodes strategy, introduced the dynamic approach for rumor influence minimization, and found a compromise method between blocking nodes and anti-rumor strategy.

The existing misinformation diffusion models and control strategies only consider the influence of users by their parent neighbors, and neglect the indirect influence of other users in their groups they are part of, that is, the effect of Echo Chamber Effect (ECE) of groups. However, some scholars choose influential nodes based on the group topology. Ghoshal et al. [14] used the group structure to statically select a group of seed nodes to publish positive information in order to effectively control misinformation as early as possible. Ni et al. [15] considered the group structure and proposed a group-based rumor blocking problem, that is, under the constraints of budget B, a group of seed users are selected as protectors from all groups, to maximize the expected number of users not affected by rumor sources. These authors only used the structure of the group/community to select seed nodes to publish positive information [14] or protect nodes [15,16], and do not consider the influence of the clustering effect of users within the group on the scope and speed of dissemination of misinformation. Based on this, this paper considers the overall impact of groups on the spread of misinformation, constructs a concrete expression of the group ECE, and studies the control strategies of misinformation in OSNs.

1.2. Contributions

In this paper, we aim to minimize the number of nodes influenced by misinformation and study the strategy of private groups disbanding in OSNs. More specifically, given a directed acyclic social network $G(V, E, C)$, a predetermined propagation source $S \subseteq V$, a budget K and an influence diffusion model $\mathscr{M}$. The goal of this paper is to find and disband the K groups from C to minimize the number of nodes ultimately activated by S. We make several contributions in this paper as follows:

First, we define different characteristics of the group in OSNs and develope a parameter function describing the group ECE.

Second, we constructe a Propagation Model of misinformation under the effect of ECE (PME), and propose a novel problem of Minimizing the Spread of Misinformation under the effect of ECE (MSME) based on PME. We use the reduction method to prove that the problem of MSME is NP-hard under PME, and it also shows that the computation of the objective function of the problem of MSME is #P-hard under PME. Moreover, we use counterexamples to prove that the objective function of the problem of MSME is a non-submodular and non-supermodular.

Third, the theoretical proof clearly shows that there is no standard optimization procedure for solving the problem of MSME, so a greedy algorithm is proposed and several group scoring heuristic algorithms are constructed.

Lastly, we use experimental simulations to determine the parameters in the misinformation dissemination, and then verify our proposed method and constructed algorithms. The experimental results reflect the effectiveness of our designed algorithm.

1.3. Organization

The rest of this paper is organized as follows. Section 2 introduces the preliminaries. Section 3 constructs the PME and describes the problem of MSME. In Section 4, a greedy algorithm and several heuristic algorithms are designed for solving the objective function. Section 5 verifies the model and algorithm through experiments, and concludes the full paper in Section 6.

2. Preliminaries

2.1. Notations

Table 1 summarizes the symbols used in the model building process of this paper and explains their meanings.

2.2. Definitions

Given a directed acyclic social network $G(V, E, C)$, where V is the node set (users), $E \subseteq V \times V$ denote an edge set (the relationships between users), $c \in C$ is a private group/community formed spontaneously by some users. Different groups have diverse characteristics in terms of group activity, group degree, and group closure, etc. Next we give a specific definition of group characteristics.

Definition 1. (Group Degree). The group degree is divided into out-degree and in-degree. The out-degree OD_i of group c_i is defined as the number of directed edges whose initial point is in N_i and the terminal point is in $V \setminus N_i$. Mathematically, it can be expressed as

$$OD_i = \sum_{v \in N_i} \left| N^{out}(v) \cap (V \setminus N_i) \right| \tag{1}$$

where N_i is the nodes within the group c_i. Similarly, the in-degree ID_i of group c_i is defined as the number of directed edges whose initial point is in $V \setminus N_i$ and the terminal point is in N_i. So, we express it as

$$ID_i = \sum_{v \in N_i} \left| N^{in}(v) \cap (V \setminus N_i) \right| \tag{2}$$

Therefore, the group degree DE_i of group c_i can be expressed as:

$$DE_i = OD_i + ID_i = \sum_{v \in N_i} \left| N^{in}(v) \cap (V \setminus N_i) \right| + \left| N^{out}(v) \cap (V \setminus N_i) \right| \tag{3}$$

Table 1
Frequently used notations.

Notation	Description
$G(V, E, C)$	A directed social network, where V is the node set, E is the edge set, C is the private group set.
P_{uv}	The probability that node u activates child neighbour node v.
$P_{i\cdot}(\cdot)$	The probability that node v is activated in c_i due to the effect of ECE of group c_i.
$\theta_v(\mathbb{C})$	The probability that node $v \in V$is activated by the S on topology $C \setminus \mathbb{C}$.
$\delta_S(\mathbb{C})$	The number of nodes that ultimately are activated by S on topology $C \setminus \mathbb{C}$.
S	The propagation source of misinformation.
N_i	Nodes within the group c_i.
C_v	Group set where node v is located.
$N^{out}(v)$	Child neighbour nodes of node v.
$N^{in}(v)$	Parent neighbour nodes of node v.

Definition 2. (Group Closure). Group closure in OSNs mainly refers to the average interaction of information between nodes within the group and nodes outside the group. So, the group closure φ_i of group c_i is defined as

$$\varphi_i = \frac{ID_i}{|N_i|} \tag{4}$$

where ID_i is the in-degree of group c_i and $|N_i|$ is the number of nodes within the group c_i.

Definition 3. (Group Density). This paper uses the clustering degree of the nodes in the group to describe the group density. The denser the group, the higher the trust between nodes, the higher the trust between nodes, the more private information nodes exchange in the group, and the higher the homogeneity of different nodes. The group density β_i of c_i is defined as

$$\beta_i = \frac{2|L_i|}{|N_i|(|N_i| - 1)} \tag{5}$$

where $|L_i|$ denotes the number of edges between nodes in c_i.

Definition 4. (Group Concern Relevance). In OSNs, users can receive different types of information, but there is a correlation between these dissimilar types of information. The intensity of correlation depends on the kind or type of information. The nodes in the group c_i pay attention to some or several types of information, and the information they pay attention to is also related to the received information I_j. In a simplified way, there is a concern correlation between the group c_i and the received information I_j. The concern relevance matrix $\mathscr{R}$ between different groups and dissimilar types of information is shown in Fig. 1. Similar to the existing work [17], three types of correlations are considered in Fig. 1.

The element $\gamma_{ij} \in [-1, 1]$ in the matrix $\mathscr{R}$, positive and negative values respectively indicate positive and negative correlations, and zero indicates irrelevant correlation. In Fig. 1, each row indicates the correlation between the information that group c_i is concerned and different types of information, and each column indicates the correlation between the information of I_j type and the concern information of different groups. In this paper, in order to conveniently calculate its influence probability, we take $\gamma_{ij} = 0$ when $\gamma_{ij} < 0$.

Definition 5. (Group Activity). The amount of group public information interaction refers to the total number of messages sent by group members to the group in a given time period. When group members have a high degree of involvement in topic discussions, the total amount of public information interaction in the group is very high; conversely, when users in the group are not interested in a certain piece of information, and almost no users participate in the discussion, the total amount of information interaction is very low. The group activity of c_i refers to the average number of interactions of public information in the group c_i in a given time interval, and group activity δ_i of c_i is defined as

$$\delta_i = \frac{PD_i}{|N_i|} \tag{6}$$

where PD_i denotes the amount of group public information interactions.

Definition 6. (Echo Chamber Effect) The concept of "Echo Chamber Effect" originated in 1956 and was mainly used in research in the political field [18]. The ECE on OSNs refers to the repletion of similar ideas or opinions in a relatively closed environment (group, community, etc.), making most people in group/community think these views or opinions are real. Different structural characteristics and user attributes make dissimilar groups show different ECEs. Therefore, the strength EC_i of the ECE in group c_i is defined as

	I_1	I_2	I_3	$\cdots$	I_n
c_1	0.4	0.3	0.9	$\cdots$	-0.5
c_2	-0.7	0	0.1	$\cdots$	0.6
c_3	0	0.3	0.6	$\cdots$	0.1
c_4	0.5	-0.4	-0.2	$\cdots$	0.8
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
c_l	0.9	0.3	0.6	$\cdots$	0.4

Fig. 1. Group Concern Relevance Matrix $\mathscr{R}$.

$$EC_i = H^i(\gamma, \beta, \delta, \varphi) = H(\gamma_i, \beta_i, \delta_i, \varphi_i) \tag{7}$$

where γ, β, δ and ϕ represent the group concern relevance, the group density, the group activity, and the group closure, respectively.

During the dissemination of information, the different characteristics of the group, such as γ, β, δ and φ, affect the strength of the group ECE. This article gives a simple EC_i expression based on the relationship between the characteristics of the group and group ECE,

$$EC_i = \frac{1}{1 + \exp -[b_1\gamma_i + b_2\beta_i + b_3\delta_i + b_4\varphi_i]} \tag{8}$$

where $b_1, b_2, b_3, b_4 \in (0, 1)$ are balance coefficients which satisfy $\sum_{i=1}^{4} b_i$.

Definition 7. (Group Disband). Group disband refers to destroying the comparatively closed environment spontaneously formed by its nodes, so that nodes in the group cannot send and receive information in a closed environment, and can only spread information through private connections (edges) between nodes. In reality, the relatively closed environment is destroyed by methods such as muting all nodes in the group or intervening with the group owner to spontaneously disband the managed group.

Over time, the topics of group concern are changeful, and the polarization intensity of information is also changing, so it is very challenging to accurately depict group ECE. This article assumes that the type of information that c_i concerns on is fixed over a given time interval Γ. In addition, the communication modes between nodes is not known in a given time interval Γ. Therefore, this article assumes that the communication modes and frequencies that existed between nodes in the past are more likely to occur in future communication.

2.3. Diffusion models

The Independent Cascade Model (ICM) [19,8,20,21] and the Linear Threshold model [22–24] are the two most classical models in the field of information diffusion. In social networks, users often forward or share information published by others, but this information usually comes from the user's close neighbors (friends), so the dissemination of information among users creates an information cascade. In addition, although people's behaviors or decisions are always strongly influenced by the number of incrementally activated users, the influence of the group ECE is also considered as an independent event. Therefore, we study the problem of MSME in this article based on the ICM. Next, we briefly review ICM here. Given a directed acyclic social network $G(V, E, C), (u, v) \in E$ has a parameter p_{uv}, which represent the probability that node u successfully activates node v. We call nodes as activated or influenced if they adopt the misinformation. Given a propagation source $S \subset V$, the misinformation dissemination process in the ICM is carried out in discrete steps $t(t = 0, 1, 2, \cdots)$, as follows:

Start with a set of initially activated nodes $S_1 = S$ at the step $t = 0$. S_1 activates its inactive child neighbour node with the probability of success $p_{(\cdot,\cdot)}$ in step $t = 1$. The newly activated nodes in step $t - 1(t > 1)$ is regarded as the new propagation nodes S_t in step t. At step t, the S_t tries to activate its inactive child neighbour nodes with a certain probability of success. When $u \in S_t$, it has one and only one chance to activate its inactive child neighbour node v with probability p_{uv} in step t. When the new propagation nodes S_t is an empty set, the information diffusion is terminated. Note that a node can only switch from inactive to active, but not in reverse direction.

Next, we use Algorithm 1 to summarize the specific process of information diffusion in the ICM.

Algorithm 1: Information diffusion process in the ICM

Input: $G(V, E, C)$, S
Output: The final activated nodes AS

$t = 0$, $A_t = S$, $AS = S$
while $A_t \neq \varnothing$ **do**
  $Med = \varnothing$
  for $u \in A_t$ **do**
    node u to activate $v \in N^{out}(V) \setminus AS$ with p_{uv}
    if node v is activated, u is added to Med and AS
  end for
  $t \leftarrow t + 1$, $A_t \leftarrow Med$
end while
return AS

---

3. Problem formulation

In this section, we first construct a propagation model of misinformation under the effect of ECE, and then give a formal statement of the problem of MSME, and finally prove the attributes of the proposed problem.

3.1. The propagation model of misinformation under the effect of ECE

In this part, based on the ICM, the propagation model of misinformation under the effect of ECE is constructed. Given a directed acyclic social network $G(V, E, C)$ and a propagation source $S \subseteq V$. Let $\mathbb{R}_t$ be the set of nodes activated in step $t(t = 0, 1, 2, \cdots)$ and $\mathbb{R}_0 = S$. At step $t(t \geqslant 1), \mathbb{R}_{t-1}$ activates inactive nodes by following two rules: firstly, $u \in \mathbb{R}_{t-1}$ has only one chance to activate its inactive child neighbour v with probability p_{uv}; secondly, if $\mathbb{R}_{t-1} \cap N_i$ is not empty, it will have only one chance to activate the inactive nodes within c_i with probability $P_{it}(\cdot)$. Repeat this process until no further nodes are activated, that is $\mathbb{R}_t = \varnothing$. It is worth pointing out that after the node is activated, the state will remain the same, until the end of the misinformation dissemination.

We take Fig. 2 as an example to illustrate the PME. The social network G in Fig. 2 consists of five nodes $\{v_1, v_2, v_3, v_4, v_5\}$ and one group $\{c_1\}$. Given a propagation source $S = \{v_2\}$, the probability $p_{uv} = 1$ of each edge (u, v), and the probability $P_{1t}(\cdot) = 1(t = 0, 1, 2, \cdots)$. The PME starts from the step $t = 0$ and triggers the $S = \{v_2\}$, namely $\mathbb{R}_0 = S = \{v_2\}$. At step $t = 1$, firstly, node v_1 and v_5 are successfully activated by their parent neighbour node v_2, that is, node v_1 and v_5 are added to $\mathbb{R}_1$; secondly, since $\mathbb{R}_0 \cap N_1 = \{v_2\}$, node v_3 is successfully activated with probability $P_{11}(\mathbb{R}_0 \cap N_1)$, at the same step, v_3 add to $\mathbb{R}_1$. At step $t = 2$, firstly, because the nodes $\{v_1, v_5\} \in \mathbb{R}_1$ has no child neighbour node, node $v_3 \in \mathbb{R}_1$ activates its child neighbour node v_4 with probability $p_{v_3 v_4}$, then, v_4 is added to $\mathbb{R}_2$; secondly, since all the nodes in c_1 are activated, no node is activated due to the effect of ECE. At step $t = 3$, since all nodes in G are activated, no node will be activated in this step, that is, $\mathbb{R}_3 = \varnothing$. Therefore, the spread of misinformation is terminated.

3.2. Problem description

It can be seen from the PME that an inactive node v in group c_i is not only directly influenced by its already activated parent neighbour nodes, but also influenced by the effect of group ECE. Therefore, this article considers the effect of group ECE in OSNs, and proposes the problem of MSME. The goal of this article is to identify and disband some groups to minimize the number of nodes ultimately are influenced by misinformation, that is, to reduce the number of nodes influenced by the effect of group ECE, so that most nodes may only receive misinformation from their activated parent neighbour nodes. The problem of MSME is defined as:

Definition 8. (MSME) Given a directed acyclic social network $G(V, E, C)$, a predetermined propagation source $S \subseteq V$, a propagation model $\mathscr{M}$, and a budget $K(1 \leqslant K \leqslant |C|)$. MSME aims to identify and disband subset $\mathbb{C}$ with $|\mathbb{C}| = K$ from C to minimize the number of nodes that ultimately are influenced by S, that is, the function $f(\mathbb{C}) = \mathbb{E}[\delta_S(\mathbb{C})]$ is the smallest, which is equal to

$$\mathbb{C}^* = \arg \min_{\mathbb{C} \subseteq C, |\mathbb{C}| = K} \mathrm{E}[\delta_S(\mathbb{C})] \tag{9}$$

where $\mathbb{E}[\cdot]$ represents an expectation operator and $\delta_S(\mathbb{C})$ represents the number of nodes finally activated by S on topology $C \setminus \mathbb{C}$.

3.3. Properties of the problem of MSME

3.3.1. Hardness results

.

Theorem 1. *The problem of MSME is NP-hard.*

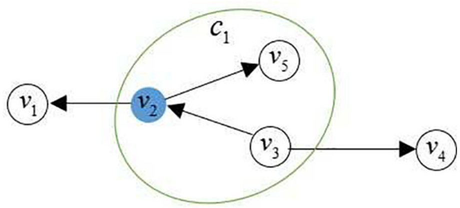


Fig. 2. An example to explain the PME.

Proof. The problem of minimizing the influence of misinformation in OSNs under the ICM [25] has been proved to be NP-hard, which is a special case of the problem of MSME when group C is empty. Therefore, the problem of MSME is obviously NP-hard under PME. □

Theorem 2. *Given a propagation source $S \subseteq V$, computing $f(\mathbb{C})$ is #P-hard under the PME.*

Proof. When $C = \emptyset$ in social network G, the problem of MSME is equivalent to the problem of Minimizing the Spread of misinformation in Social network under the ICM (MS2I). Now that we already know that the influence spread computation problem in social network under ICM is #P-hard [20], since the problem of MS2I is a special case of the problem of MSME, so the influence spread computation problem under PME is also #P-hard, that is, computing $f(\mathbb{C})$ is #P-hard under the PME. □

3.3.2. Modularity of objective function

The objective function of the problem of MS2I is non-submodular [20]. The problem of MSME can be reduced to the problem of MS2I, so unfortunately, the objective function of the problem of MSME is also non-submodular. Next, we will use a simple counterexample to prove that the $f(\cdot)$ is neither submodular nor supermodular.

Theorem 3. *$f(\cdot)$ is not submodular under the PME*

Proof. We use Fig. 3 to construct a counterexample to prove that the $f(\cdot)$ is non submodular. Let $A1 = \emptyset, A2 = \{c_1\}$ and $c = \{c_3\}$, we have $f(A1) = 9.79, f(A2) = 8.849, f(A1 \cup c) = 8.71$ and $f(A2 \cup c) = 7.868$. That is,

$f(A1 \cup c) - f(A1) < f(A2 \cup c) - f(A2)$. where $A1 \subseteq A2 \subseteq C$ and $c \in C \setminus A2$. Therefore, $f(\cdot)$ is not submodular. □

Theorem 4. $f(\cdot)$ is not supermodular under the PME.

Proof. Similarly, we use Fig. 3 to construct a counterexample to prove that the $f(\cdot)$ is non supermodular. Considering $A1 = \emptyset,\ A2 = \{c_2\}$ and $c = \{c_3\}$, we have $f(A1) = 9.79,\ f(A2) = 8.403, f(A1 \cup c) = 8.71$ and $f(A2 \cup c) = 6.609$. Thus, we have $f(A1 \cup c) - f(A1) > f(A2 \cup c) - f(A2)$ for some $A1 \subseteq A2 \subseteq C$ and $c \in C \setminus A2$, which proves that $f(\cdot)$ is not supermodular. □

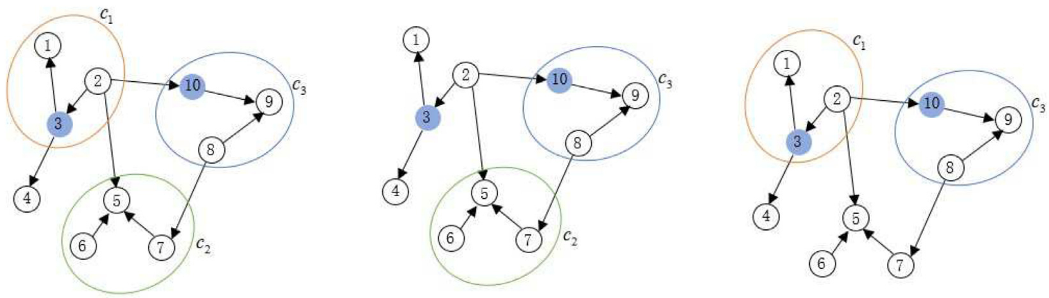
From the above discussion, we can see that the objective function $f(\cdot)$ is neither submodular nor supermodular. Next, we will briefly analyse the reason why $f(\cdot)$ has no submodular and supermodular.

In Fig. 4a, the loss of misinformation propagation value when disbanding group c_1 is composed of $\phi_{C_1}^{V_1}(c_1), \phi_{C_1}^{V_2}(c_1), \phi_{C_1}^{V_3}(c_1)$ and $\phi_{C_1}^{V_4}(c_1)$, where $\phi_{C_1}^{V_1}(c_1)$ represents the total number of reduction values of the probability that the nodes in V_1 are activated by S when c_1 is disbanded on topology C_1. Similarly, when c_1 is disbanded in Fig. 4b, the loss of misinformation propagation value is also composed $\phi_{C_2}^{V_1}(c_1), \phi_{C_2}^{V_2}(c_1), \phi_{C_2}^{V_3}(c_1)$ and $\phi_{C_2}^{V_4}(c_1)$. Considering $C_2 \subset C_1$ only considering the effect of group c_1 on the V_3, we have $\phi_{C_1}^{V_3}(c_1) \leqslant \phi_{C_2}^{V_3}(c_1)$. But it can be seen that V_3 is influenced not only by group c_1 but also by group c_2. When group c_1 is disbanded, the strength of the ECE of group c_2 is also weakened, which in turn increases $\phi_{C_1}^{V_3}(c_1)$, so we cannot determine the value of $\phi_{C_1}^{V_3}(c_1)$ and $\phi_{C_2}^{V_3}(c_1)$. Therefore, it can be seen from here that the objective function $f(\cdot)$ is neither submodular nor supermodular.

4. Solution methods

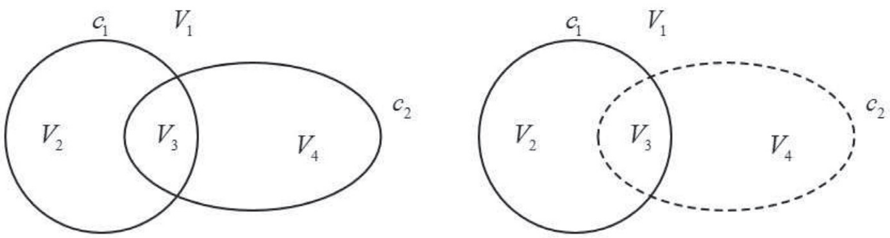
In social network $G(V, E, C)$, let $\theta_v(C \setminus \mathbb{C}, S)$ denote the probability that node $v \in V \setminus S$ is activated by the S on topology $C \setminus \mathbb{C}$, we abbreviate it as $\theta_v(\mathbb{C})$; p_{uv} represent the probability that node u activates child neighbour node v and $P_{i\cdot}(\cdot)$ represent the probability that node v is activated in c_i due to the influence of ECE of group c_i.

Next, we will give the calculation process of probability $\theta_v(\mathbb{C})$ with which node $v \in V \setminus S$ is activated by S on the topology $C \setminus \mathbb{C}$ under the PME. Given a graph $G(V, E, C)$ and a time limit Γ, we build a Γ-layer graph G^Γ to calculate the probability $\theta_v(\mathbb{C})$. G^Γ is on the nodes of $\Gamma \cdot |V|$, v^t and $c^t (t \leqslant \Gamma)$ are copies of node v and group c of graph G, respectively. Let $\theta_v^t(\mathbb{C})$ be the probability that node v is activated by S in the first t layer, $\theta_{v^t}(\mathbb{C})$ and $\triangle\theta_{v^t}(\mathbb{C})$ is the initial probability and the increased activation probability of node v^t in graph G^t, respectively. Let Λ^t be the nodes with $\theta_{v^t}(\mathbb{C}) > 0$ in G^t. The initial probability of node v^{t+1} being activated in graph G^{t+1} is the increased probability of node v^t being activated in graph G^t. Let the $\theta_{v^0}(\mathbb{C}) = 0$ of $v^0 \in V^0$ in the graph $G^0, \triangle\theta_{v^0}(\mathbb{C}) = 1(v^0 \in S)$ and $\triangle\theta_{v^0}(\mathbb{C}) = 0(v^0 \notin S)$. We give $\theta_v^0(\mathbb{C}) = 1(v \in S)$ and $\theta_v^0(\mathbb{C}) = 0(v \notin S)$. In graph G^t, the increased probability of node v^t being activated is.



(a) A simple social network. (b) Disband $\mathbb{C} = \{c_1\}$. (c) Disband $\mathbb{C} = \{c_2\}$.

Fig. 3. A counterexample. We give the influence probability $p_{uv} = 1$ of each edge (u, v), and the influence probability $P_{i\cdot}(\cdot) = 0.5$ of the ECE of group c_i. It is worth noting that the blue nodes $\{3, 10\}$ are the propagation source S.



(a) Network topology C_1 . (b) Network topology C_2 .

Fig. 4. An example to explain why the objective function is non-submodular and non-supermodular. The topology of Fig. 4a and Fig. 4b are $C_1 = \{c_1, c_2\}$ and $C_2 = \{c_1\}$, respectively. The nodes in Fig. 4a are divided into four parts $\{V_1, V_2, V_3, V_4\}$, in order to compare Fig. 4b with Fig. 4a, we use a dashed circle in Fig. 4b to indicate the disbande.d group c_2.

$$\Delta\theta_{v^t}(\mathbb{C}) = \left[1 - \Pi_{c_i^t \in C_v}\left[1 - P_{it}\left(\sum_{d^t \in N_i^t} \theta_{d^t}(\mathbb{C})\right)\right] \cdot \prod_{u^t \in N^{in}(v^t) \cap \Lambda^t} (1 - \theta_{u^t}(\mathbb{C}) \cdot p_{u^t v^t})\right] \left(1 - \theta_v^{t-1}(\mathbb{C})\right)$$

where C_v represents the groups where node v is located. Then, the probability $\theta_v(\mathbb{C})$ is.

$$\theta_v^t(\mathbb{C}) = 1 - \left(1 - \theta_v^{t-1}(\mathbb{C})\right)\left(1 - \triangle\theta_{v^t}(\mathbb{C})\right)$$

Therefore, the $\theta_v(\mathbb{C})$ of node $v \in V$ can be expressed as $\theta_v(\mathbb{C}) = \theta_v^{\Gamma}(\mathbb{C})$. Then, the objective function can be equivalent to

$$\mathbb{C}^* = \arg \min_{\mathbb{C} \subseteq C, |\mathbb{C}| = K} \sum_{v \in V} \theta_v(\mathbb{C}) \tag{10}$$

From the Eq. (10), we convert the expected value of minimizing the total number of activated nodes into minimizing the probability of each node being activated by S.

Since there is currently no standard optimization procedure that can be adopted to obtain an approximate guaranteed solution of a non-submodular function or a non-supermodular function, in this section we first propose a greedy algorithm to solve the problem of MSME based on the greedy strategy, and then construct several heuristic algorithms to solve this problem according to the attributes of the group in the OSNs.

4.1. Greedy strategy method

This section, based on the hill climbing algorithm [26], develops a Greedy Algorithm (GA) for group selection, which starts with an empty set and iteratively selects the number of groups until a given K is reached. The idea of GA is to select the group that causes the largest reduction of $f(\cdot)$ to be disbanded or shielded each time, that is, to disband the group with the largest of the effect of group ECE each time. Although the GA of the problem of MSME does not provide the accuracy guarantee of the optimal solution of $1 - 1/e$, it also gives a comparatively good solution in the real-world social networks. The specific steps of the GA are presented in Algorithm 2.

Algorithm 2: Greedy Algorithm (GA)

Input: $G(V, E, C)$, S, K
Output: $< \mathbb{C}^*, f(\mathbb{C}^*) >$
1: $\mathbb{C} = \varnothing$
2: **while** $|\mathbb{C}| < K$ **do**
3: $\quad c \leftarrow \arg\min_{c \in C \setminus \mathbb{C},} (GSVM(\mathbb{C} \cup \{c\}, S) - GSVM(\mathbb{C}, S))$
4: $\quad \mathbb{C} \leftarrow \mathbb{C} \cup \{c\}$
5: **end while**
6: $\mathbb{C}^* \leftarrow \mathbb{C}, f(\mathbb{C}^*) \leftarrow GSVM(\mathbb{C})$
7: **return** $< \mathbb{C}^*, f(\mathbb{C}^*) >$

In Algorithm 3, we first initialise $\mathbb{C}$ as an empty set, and then select the group c that minimize the total probability of all nodes being activated to join $\mathbb{C}$ in each iteration, and stop when K groups are selected. It can be seen from the third line of Algorithm 2 that when c is selected for each iteration, the spread value of misinformation needs to be recalculated, so we give specific steps for calculating the spread value of misinformation in Algorithm 3.

Algorithm 3: Get the Spread Value of Misinformation (GSVM)

Input: $G(V, E, C)$, S, $\mathbb{C}$, Γ
Output: $f(\mathbb{C})$
1: $\overline{G}\left(V, E, \overline{C}\right) \leftarrow G(V, E, C \setminus \mathbb{C})$
2: Generate the Γ-layer graph $\overline{G}^{\Gamma}$
3: Initialize $\theta_{v^0}(\mathbb{C})$, $\triangle\theta_{v^0}(\mathbb{C})$, $\theta_v^0(\mathbb{C})$ for each $v^0 \in V^0$
4: $t = 1$, $\theta_{v^1}(\mathbb{C}) \leftarrow \triangle\theta_{v^0}(\mathbb{C})$, Λ^1, $\theta_v^1(\mathbb{C})$
5: **while** $t \leqslant \Gamma$ and $\Lambda^t \neq \varnothing$ **do**
6: $\quad$ Calculate the $\triangle\theta_{v^t}(\mathbb{C})$ for each node based on Eq. (6)
7: $\quad t+ = 1$, $\theta_{v^t}(\mathbb{C}) \leftarrow \triangle\theta_{v^{t-1}}(\mathbb{C})$, Λ^t, $\theta_v^t(\mathbb{C})$
8: **end while**
9: Calculate the $\theta_v(\mathbb{C})$ for $v \in V$
10: $f(\mathbb{C}) = \sum_{v \in V} \theta_v(\mathbb{C})$
11: **return** $f(\mathbb{C})$

In Algorithm 3, the graph $\overline{G}$ is obtained by updating the topology of G according to $\mathbb{C}$, and then the graph $\overline{G}$ is copied Γ times to get the graph set $\overline{G}^{\Gamma}$. Secondly, the $\overline{G}^{\Gamma}$ is used to calculate the $\theta_v(\mathbb{C})$ of the node $v \in V$ being activated by S on topology $C \setminus \mathbb{C}$, and finally, the $f(\mathbb{C})$ of the total probability of all nodes being activated by S is obtained.

4.2. Heuristic methods

In real world, OSNs are very large and contain many groups, so it takes a lot of memory and time to solve the objective function using GA. Therefore, in this section, we constructively propose several heuristic algorithms for solving the objective function based on group attributes such as the group degree, group density and group closure, etc. Compared with GA, these heuristic algorithms have the advantages of shorter running time and less memory usage. Next, we'll outline the ideas of these heuristic algorithms.

- Scoring of Group Degree Algorithm (SGDA). The group degree represents the sum of the misinformation channels that group c_i receives from outside the group and sends misinformation to the outside. The larger the ID_i, the wider the range of misinformation it receives, and the smaller the polarization of the thought of the members in the group, that is, the weaker the effect of group ECE. The larger the OD_i, the stronger the ability of the group to spread misinformation to the outside world, that is, the stronger the ability to spread the group ECE to nodes outside the group. Therefore, the score $gd(c_i)$ of $c_i \in C$ in SGDA is defined as

$$gd(c_i) = \frac{OD_i}{ID_i} \tag{11}$$

SGDA selects the group with the highest $gd(c_i)$ to shield or disband each time, so as to minimize the number of users who ultimately are activated by S.

- Scoring of Group Closure Algorithm (SGCA). The group closure φ_i represents the average misinformation interaction between the N_i and $V \setminus N_i$. The less misinformation interaction a node $v \in N_i$ has with nodes outside the group, the less misinformation it obtains from outside the group, and the easier it is to be influenced by the misinformation in the group. When there is more misinformation interaction between inside and outside the group, the more misinformation is obtained by the nodes in the group, then the users in the group will compare the misinformation received in the group with the misinformation obtained from outside the group and then decide whether to accept it or not. Therefore, the score $gc(c_i)$ of $c_i \in C$ in SGCA is defined as

$$gc(c_i) = \frac{\varphi_i}{OD_i} \tag{12}$$

So, we select the group with the lowest $gc(c_i)$ to disband each time.
- Scoring of Group Density Algorithm (SGEA). The group density β_i indicates the degree of aggregation between nodes within group c_i. The greater the β_i, the more private misinformation interaction between nodes in the group, and the weaker the effect of the ECE. When the group is denser, the more private exchanges among users in the group, the higher the trust between nodes, and the higher the homogeneity of different nodes. Therefore, the score $de(c_i)$ of $c_i \in C$ in SDEA is defined as

$$de(c_i) = |\beta_{max} + \beta_{min} - 2\beta_i| \tag{13}$$

where β_{max} and β_{min} represent the maximum and minimum values of β_i, respectively. We iteratively select K groups with the lowest $de(c_i)$ to disband or shield in order to minimize the number of users who ultimately accept misinformation.
- Random Algorithm (RDA). The main idea of RDA is to randomly and uniformly select K groups from C to disband or shield, so as to minimize the number of users who ultimately are activated by S.

5. Experiments

Here we first give a descriptive explanation of the real world datasets, then give the specific equation for the influence probability of misinformation and present the parameter values in the process of misinformation dissemination, and finally analyse and discuss the main experimental results.

5.1. Datasets

This article focuses on four real-world datasets that contain group structures, as shown in Table 2. Email-EU [27] dataset was generated from email data of large European research institutions. Each edge (u, v) indicates that user u sent at least one email to user v. The dataset also contains the "real-life" group membership of the users and each user is affiliated with one of the institute's 42 departments. YouTube [28] dataset includes the network link structure on YouTube website and the information of each group member. The node v represents a user member and the edge (u, v) indicates that user u has shared video with user v on the YouTube website. $c \in C$ represents an interest or private group composed of multiple users together. Ning [29] dataset is a snapshot of the social network friendship and group affiliation network collected in September 2012, where nodes represent users and the edge (u, v) indicates that there is a friendly relationship between user u and v in the social media Ning. DBLP [30] dataset denotes a complete list of computer science research papers, where nodes represent authors, and when two authors have published at least one paper, they are connected as the edge, and the authors who published papers in a certain journal or conference form a group.

5.2. The experimental setup

5.2.1. Influence probability

We can see that the four datasets given in Section 5.1 all lack the influence probability between node pairs, so we give two methods to calculate the influence probability of misinformation: the first method is that the influence probability is a fixed value, we set $p_{uv} = 0.1$; the second method is the trivalent model distribution probability ($P = TRI$), which is to randomly select a value from $\{0.2, 0.05, 0.01\}$ as the influence probability. The probability of node $v \in N_i$ is activated due to the effect of ECE of the group c_i, is defined it based on the Ising model and Boltzmann distribution. So, the influence probability $P_{it}(x)$ due to the effect of the ECE of the group c_i in t step is

Table 2
Dataset statistical description.

Dataset	Relationships	#Nodes	#Edges	#Groups
Ning	Social	10 K	76 K	80
Emails-EU	Social	1 K	26 K	41
YouTube	Social	2 K	33 K	3065
DBLP	Collaboration	9 K	28 K	2706

$$P_{it}(x) = \frac{EC \cdot \exp\left(-\frac{1}{x}\right)}{\exp\left(-\frac{1}{1-x}\right) + \exp\left(-\frac{1}{x}\right)} \tag{14}$$

where x is the ratio of newly activated nodes in c_i to N_i in step t, EC is the parameter that affects the strength of the ECE and $0 \leqslant EC \leqslant 1$. Since each group has diverse characteristics and attributes, different groups have dissimilar effects on misinformation propagation, that is to say, different groups have different EC values. Therefore, the influence probability $P_{it}(I_{it})$ can be expressed as

$$P_{it}(I_{it}) = \frac{EC_i \cdot \exp\left(-\frac{|N_i|}{|I_{it}|}\right)}{\exp\left(-\frac{|N_i|}{|N_i|-|I_{it}|}\right) + \exp\left(-\frac{|N_i|}{|I_{it}|}\right)} \tag{15}$$

where $|N_i|$ is the number of nodes within the group c_i and $I_{it} = \sum_{v^t} \theta_{v^t}(\mathbb{C})$ is the sum of the increased probability of being activated for the nodes in group c_i at step t.

5.2.2. Parameter setting

To study the control strategy of misinformation in the problem of MSME, it is necessary to clarify the topics of concern in various groups in OSNs, but in the real world OSN, the specific topic types concerned by the groups are very difficult to obtain, but also the types of misinformation received each time are different. Therefore, a uniform and random selection between $[0, 1]$ is generated as the γ_{ij} of the group c_i. Besides, in OSNs, it is also difficult to obtain the amount of public misinformation interaction of group users in a given time Γ, so this article does not consider the amount of public misinformation interaction in the group, that is, the coefficient b_3 is equal to 0. In the following experiments, we set $b_1 = 0.3, b_2 = 0.05, b_4 = 0.65$ and $\Gamma = 5$, and give specific experiments in Section 5.2.3. to explain why these parameters are set to these values. We randomly and uniformly select 1% of all nodes in each dataset as S. Because some users have accepted it when misinformation is detected, We treat the child neighbor nodes of S as activated nodes at step $t = 0$. Finally, considering that there are fewer groups in Email-EU and Ning datasets, while there are more groups in DBLP and YouTube datasets, we select 50% of the total group in Email-EU and Ning datasets to disband, and 10% of the total group in DBLP and YouTube datasets to disband or shield.

5.2.3. Experimental results

.

1. Experimental results of parameter settings.
   Experiments were carried out on four real world datasets to determine the optimal values of parameters b_1, b_2 and b_4. Since the parameters b_1, b_2 and b_4 are balance coefficients, that is, $b_1 + b_2 + b_4 = 1$, one of the parameters can be expressed by the others two parameters. We determine the value range of b_2 as $\{0.05, 0.1, 0.2, 0.3\}$, and then dynamically determine the values of b_1 and b_4 according to the value range of b_2. Our experimental results are shown in Fig. 5.
   Fig. 5 shows the change of the spread value of misinformation in four datasets when different parameter values are selected. In Fig. 5a and 5b, we can see that with the same influence probability and different datasets, different parameters affect the trend of misinformation propagation differently. From the overall perspective of Fig. 5, when the parameter b_2 is given, the spread value of misinformation gradually decreases with the gradual increase of the parameter b_1. Similarly, the spread value of misinformation decreases with the gradual increase of parameter b_2 when the parameter b_1 is given. It is observed that under different datasets and different influence probabilities, the spread value of misinformation is the minimum when $b_1 = 0.6, b_2 = 0.3$, and when $b_1 = 0.3, b_2 = 0.05$, the misinformation propagation value is the largest. Therefore, in order to better describe the ECE and propose a more effective method to solve the problem of MSME, we set the parameter values $b_1 = 0.3, b_2 = 0.05, b_4 = 0.65$.
   Next, based on the given values of parameters b_1, b_2 and b_4, we will simulate on the four datasets with $P = 0.1$ and $P = TRI$, respectively, to determine the value of the misinformation propagation time limit Γ. The specific experimental simulation results are shown in Fig. 6.
   Under the different datasets and influence probabilities in Fig. 6, the propagation trend of misinformation is basically the same. With the gradual increase of misinformation propagation time, the increment of misinformation propagation value gradually decreases. On the whole, the spread value of misinformation tends to be stable when the propagation time of misinformation $\Gamma \geqslant 5$. Therefore, in this paper, we set the propagation time of misinformation $\Gamma = 5$. Since the given S contains two levels of nodes, the spread value of misinformation basically does not change when $\Gamma = 6$.
2. Experimental results of algorithms comparison. In this part, we evaluate our developed algorithms on four real world datasets by changing the number of disbanded groups. The experimental results of the Ning and Email datasets are shown in Table 3 and Table 4, respectively, while the experimental results of the DBLP and YouTube datasets are shown in Fig. 7. The horizontal axis indicates the number of disbanded groups, and the vertical axis represents the spread value of misinformation when different numbers of groups are disbanded. The baseline represents the spread value of misinformation when no group is disbanded and the NOECE indicates the spread value of misinformation without considering the ECE. We define the overall influence of the group as the difference between baseline and NOECE. When the spread value of

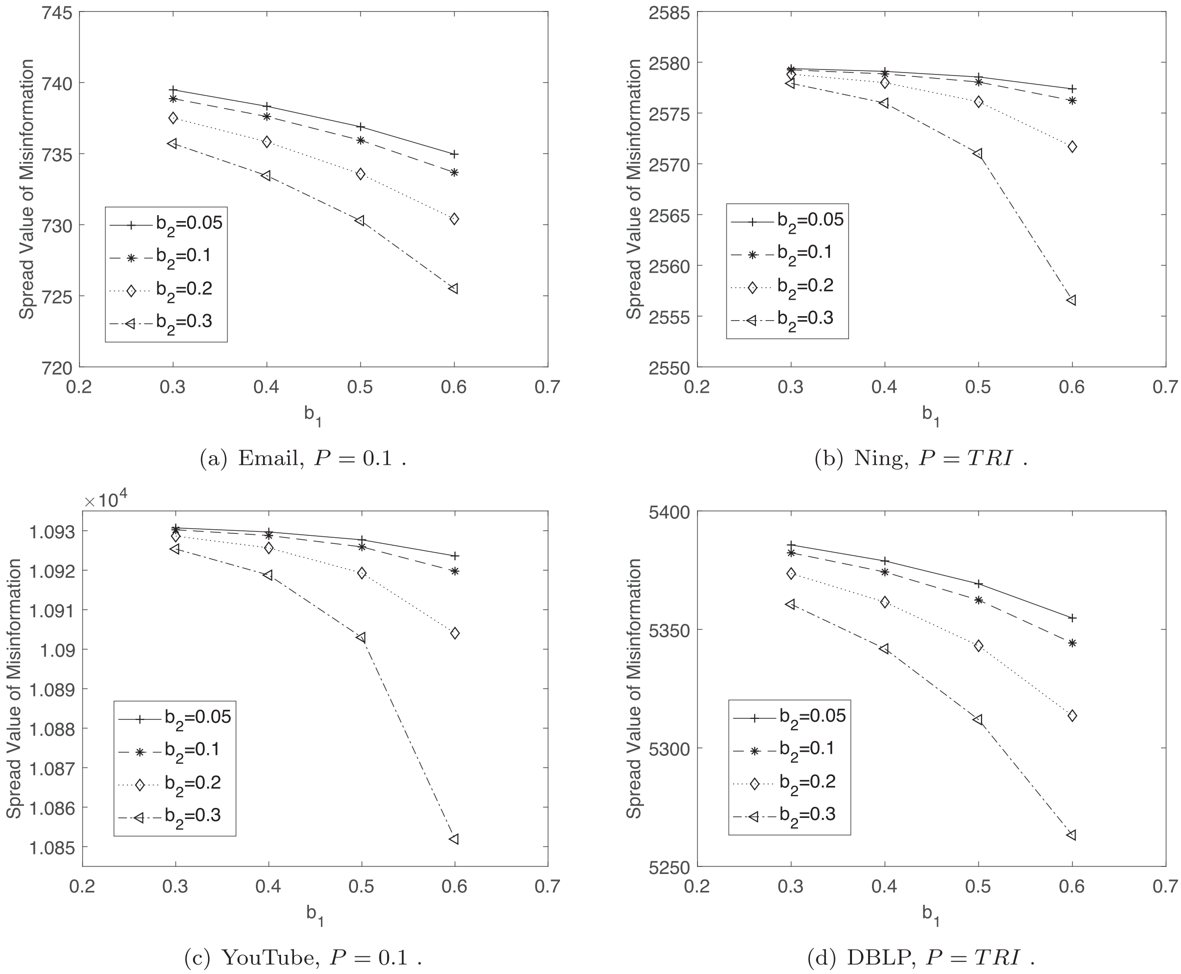
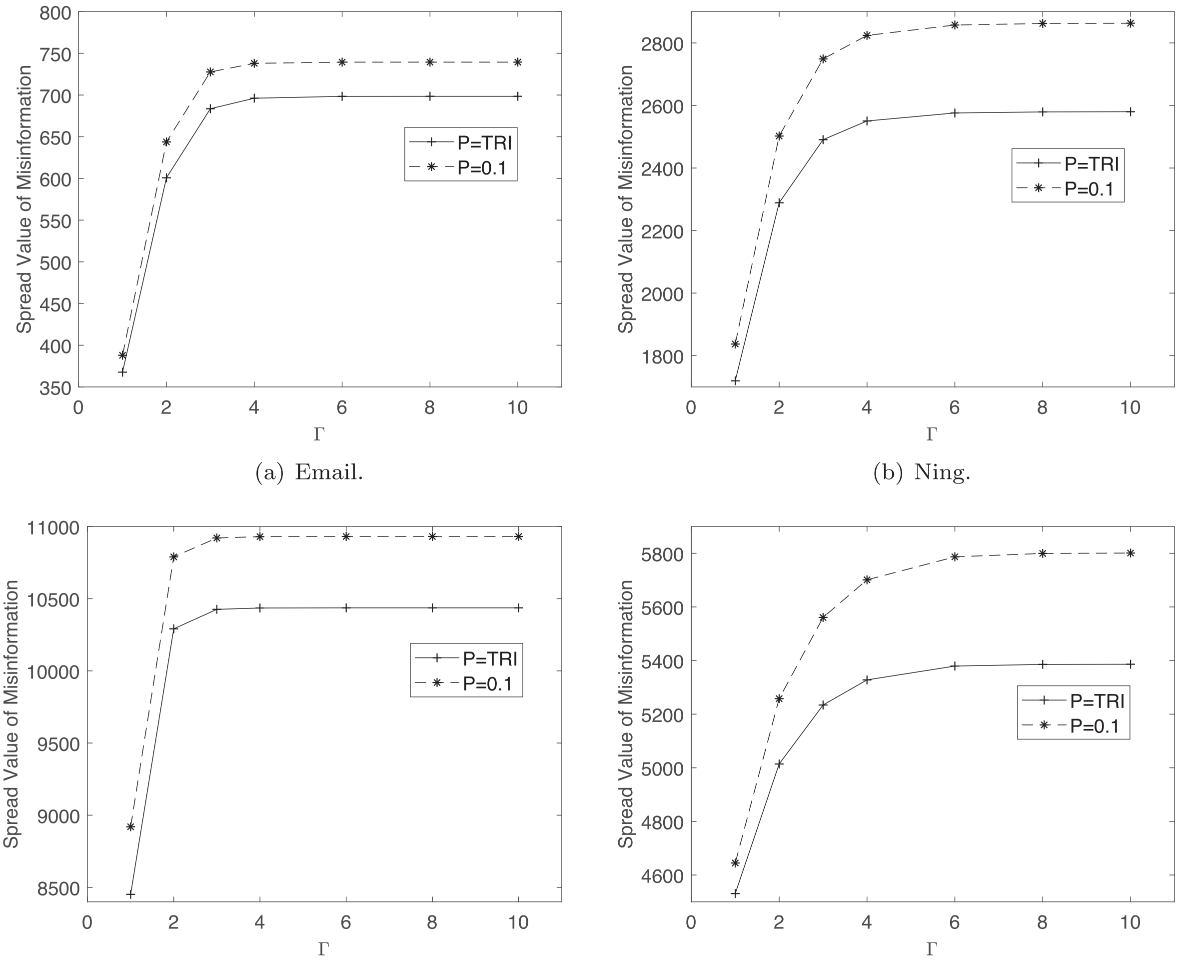


Fig. 5. Spread value of misinformation under different parameters.

misinformation is closer to the NOECE, the overall influence of the group is smaller, and vice versa, the overall influence of the group is greater. From the experimental results of the four real world datasets, we can draw the following conclusions. For different algorithms, as the number of disbanded groups increases, the spread value of misinformation is gradually decreases. For different influence probabilities in the same dataset, the same algorithm has basically the same effect in controlling misinformation.

The ability of the same algorithm to limit the spread of misinformation under different datasets is quite different. In Table 3 and Table 4, the performance of SGCA to control misinformation is significantly better than SGEA. However, in Fig. 7c and 7d, the ability of SGCA to control the spread of misinformation is basically the same as other heuristic algorithms, and even worse than SGEA. From Table 3, Table 4 and Fig. 7 as a whole, the GA has the best performance in controlling the spread of misinformation, followed by the SGEA, and the RDA has the worst performance in controlling the spread of misinformation. In Fig. 7a, we can find that when the number of disbanded groups is K, the spread value of misinformation under the GA is closer to the NOECE, which means that the overall influence of the group is smaller at this time. In Fig. 7b, 7c and 7d show similar effects, reflecting the performance of GA in controlling misinformation.

Moreover, we can see that the performance of the heuristic algorithm is unstable under different datasets or influence probabilities, and its effect of controlling the spread of misinformation is general. From Table 3 and Table 4, it can be found that the ability of SGDA to control misinformation is basically the same as that of the RDA, and even worse than the RDA. The reason for this situation may be that the number of groups in the Emails-EU and Ning datasets is relatively small and the number of disbanded groups is relatively large, which in turn leads to the invalidation of the SGDA evaluation index. For all algorithms, in the YouTube dataset, as the number of disbanded groups increases, the reduction in the spreading value of misinformation is gradually decreasing, and there is no cliff-like decline. However, when relatively



(a) Email. (b) Ning. (c) YouTube. (d) DBLP.

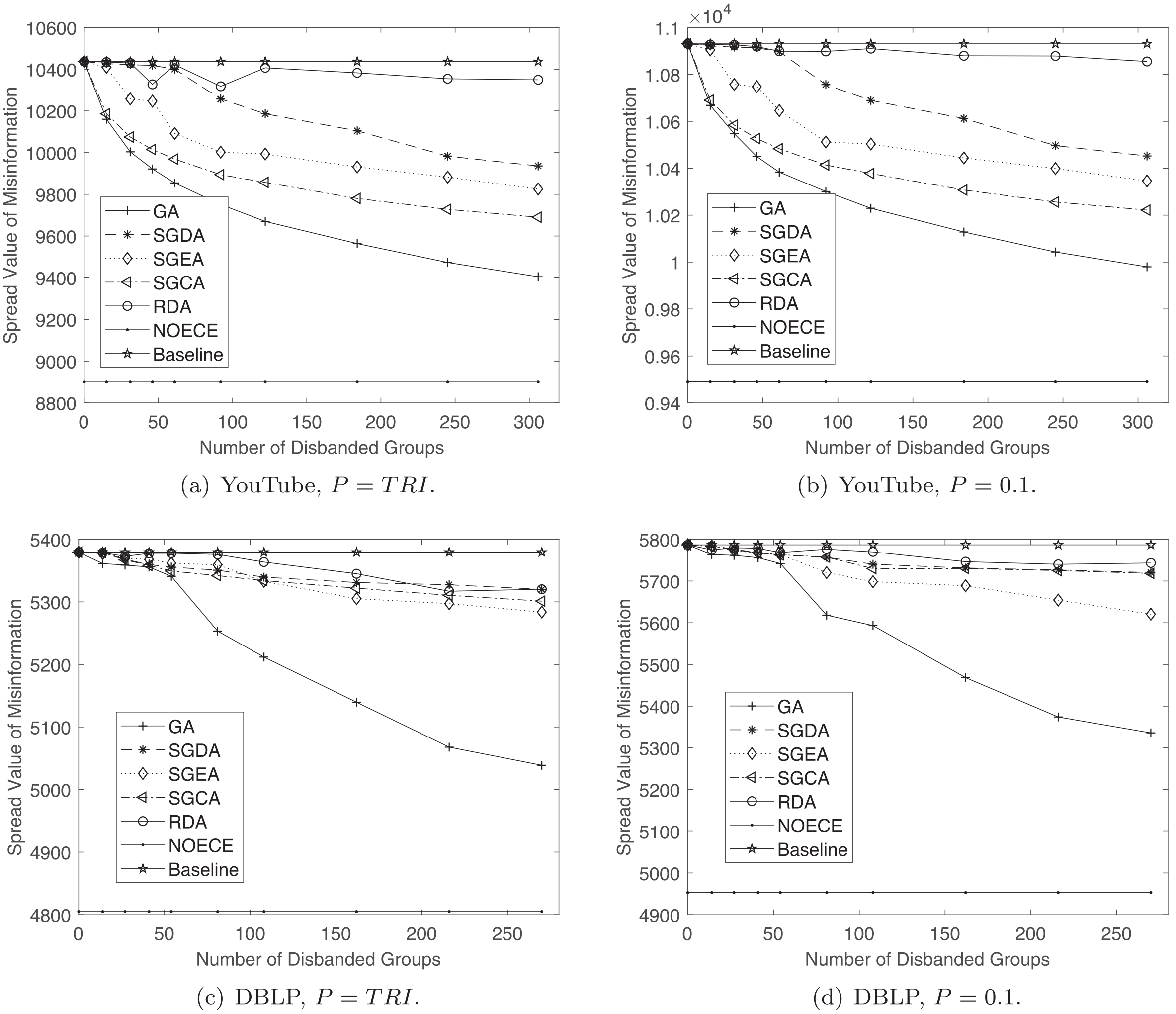
Fig. 6. The spread value of misinformation vs. The propagation time Γ.

Table 3
The spread value of misinformation of different algorithms on Emails-EU dataset.

Method	TRI					FIX				
$\|C\|$	GA	SGDA	SGEA	SGCA	RDA	GA	SGDA	SGEA	SGCA	RDA
0	698.4	698.4	698.4	698.4	698.4	739.4	739.4	739.4	739.4	739.4
2	689.8	697.3	692.7	692.8	697.2	727.6	738.3	732.8	728.8	738.1
5	679.3	696.3	685.9	681.6	693.7	715.7	737.2	723.9	718.2	736.1
10	671.3	689.7	680.2	673.9	685.0	705.7	730.4	717.3	709.3	730.4
20	655.8	673.1	660.5	657.3	679.0	687.8	710.4	695.1	691.2	718.3

Table 4
The spread value of misinformation of different algorithms on Ning dataset.

Method	TRI					FIX				
$\|C\|$	GA	SGDA	SGEA	SGCA	RDA	GA	SGDA	SGEA	SGCA	RDA
0	2576.0	2576.0	2576.0	2576.0	2576.0	2857.3	2857.3	2857.3	2857.3	2857.3
5	2536.2	2573.6	2571.0	2545.9	2566.6	2808.4	2854.9	2852.4	2813.7	2845.5
10	2514.4	2570.6	2557.5	2525.4	2560.6	2766.7	2852.1	2835.6	2787.9	2839.5
20	2496.4	2549.4	2518.0	2506.2	2537.3	2744.3	2830.5	2789.8	2763.4	2821.2
40	2449.8	2503.2	2478.9	2462.0	2521.3	2708.5	2775.9	2740.5	2718.2	2759.1



(a) YouTube, $P = TRI$.
(b) YouTube, $P = 0.1$.
(c) DBLP, $P = TRI$.
(d) DBLP, $P = 0.1$.

Fig. 7. The misinformation spread value of all method on DBLP and YouTube.

few groups are disbanded in the DBLP dataset, the reduction of misinformation propagation value is almost unchanged. When the number of blocked groups reaches a certain threshold, the misinformation propagation value drops precipitously.

6. Conclusion and future work

In this article, a novel problem of MSME is constructed based on the PME to minimize the spread of misinformation. We prove that the problem of MSME is NP-hard, and then prove that the objective function of the problem of MSME is both non-submodular and non-supermodular. We define the different characteristics of the groups in OSNs, and based on this we construct a parameter function describing the group ECE, and given specific parameter values in real world datasets form experiments. It is worth noting that GA and heuristic algorithms are constructed to solve the problem of MSME. Finally, we validate our proposed model and algorithms on four real world datasets using different influence probabilities. For future research, we want to continue to study new control strategies for misinformation based on the group ECE.

CRediT authorship contribution statement

Jianming Zhu: Conceptualization, Methodology, Writing - review & editing. **Peikun Ni:** Software, Formal analysis, Writing - original draft, Writing - review & editing. **Guoqing Wang:** Investigation. **Yuan Li:** Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Liu, L. Nie, X. Wang, Q. Tian, B. Chen, Online data organizer: micro-video categorization by structure-guided multimodal dictionary learning, IEEE Trans. Image Process. 28 (3) (2019) 1235–1247.
[2] L. Nie, X. Song, T.-S. Chua, Learning from multiple social networks, Synthesis Lectures Inf. Concepts Retrieval Services 8 (2) (2016) 1–120.
[3] B. Wang, G. Chen, L. Fu, L. Song, X. Wang, Drimux: dynamic rumor influence minimization with user experience in social networks, IEEE Trans. Knowl. Data Eng. 29 (10) (2017) 2168–2181.
[4] L. Nie, L. Zhang, M. Wang, R. Hong, A. Farseev, T.-S. Chua, Learning user attributes via mobile social multimedia analytics, ACM Trans. Intell. Syst. Technol. 8(3). .
[5] C. Sun, H. Liu, M. Liu, Z. Ren, T. Gan, L. Nie, Lara: attribute-to-feature adversarial learning for new-item recommendation, Houston, TX, United states, 2020, pp. 582 – 590. .
[6] M. Zhang, X. Wang, X. Qin, A rumor blocking method based on tolerance of users, Shenyang, China, 2019, pp. 462 – 468. .
[7] J. Zhu, P. Ni, G. Wang, Activity minimization of misinformation influence in online social networks, IEEE Trans. Comput. Soc. Syst. 7 (4) (2020) 897–906.
[8] M. Kimura, K. Saito, H. Motoda, Minimizing the spread of contamination by blocking links in a network, in: Proceedings of the 23rd National Conference on Artificial Intelligence – Volume 2, AAAI'08, AAAI Press, 2008, p. 1175–1180. .
[9] F. Wang, J. She, Y. Ohyama, W. Jiang, G. Min, G. Wang, M. Wu, Maximizing positive influence in competitive social networks: A trust-based solution, Inf. Sci. 546 (2021) 559–572.
[10] N. Arazkhani, M.R. Meybodi, A. Rezvanian, Influence blocking maximization in social network using centrality measures, Tehran, Iran, 2019, pp. 492 – 497. .
[11] L. Yang, Z. Li, A. Giua, Containment of rumor spread in complex social networks, Inf. Sci. 506 (2020) 113–130.
[12] A.I.E. Hosni, K. Li, S. Ahmad, Minimizing rumor influence in multiplex online social networks based on human individual and social behaviors, Inf. Sci. 512 (2020) 1458–1480.
[13] A.I.E. Hosni, K. Li, S. Ahmad, Darim: Dynamic approach for rumor influence minimization in online social networks, vol. 11954, LNCS, Sydney, NSW, Australia, 2019, pp. 619–630.
[14] A.K. Ghoshal, N. Das, S. Das, Influence of community structure on misinformation containment in online social networks, Knowl.-Based Syst. 213. .
[15] Q. Ni, J. Guo, C. Huang, W. Wu, Community-based rumor blocking maximization in social networks: Algorithms and analysis, Theoret. Comput. Sci. 840 (2020) 257–269.
[16] J. Zheng, L. Pan, Least cost rumor community blocking optimization in social networks, Shanghai, China, 2018. .
[17] S.A. Myers, J. Leskovec, Clash of the contagions: Cooperation and competition in information diffusion, Brussels, Belgium, 2012, pp. 539 – 548. .
[18] D. A, An economic theory of democracy, Public Choice 19 (1957) 111–115. .
[19] W. Ju, L. Chen, B. Li, W. Liu, J. Sheng, Y. Wang, A new algorithm for positive influence maximization in signed networks, Inf. Sci. 512 (2020) 1571–1591.
[20] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, New York, NY, USA, 2003, p. 137–146. .
[21] Y. Mashayekhi, M.R. Meybodi, A. Rezvanian, Weighted estimation of information diffusion probabilities for independent cascade model, in: 2018 4th International Conference on Web Research (ICWR), 2018, pp. 63–69.
[22] N.P. Nguyen, G. Yan, M.T. Thai, S. Eidenbenz, Containment of misinformation spread in online social networks, in: Proceedings of the 4th Annual ACM Web Science Conference, WebSci '12, Association for Computing Machinery, New York, NY, USA, 2012, p. 213–222. .
[23] X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, Anaheim, CA, United states, 2012, pp. 463–474.
[24] F. Zhou, J.R. Jiao, B. Lei, A linear threshold-hurdle model for product adoption prediction incorporating social network effects, Inf. Sci. 307 (2015) 95–109.
[25] C. Budak, D. Agrawal, A. El Abbadi, Limiting the spread of misinformation in social networks, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 665–674.
[26] G. Corneuejols, M.L. Fisher, G.L. Nemhauser, Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms, Manage. Sci. 23 (8) (1977) 789–810.
[27] J. Leskovec, J. Kleinberg, C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Trans. Knowl. Discov. Data 1 (1) (2007) 2–es. .
[28] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, New York, NY, USA, 2007, p. 29–42. .
[29] O. Lesser, L. Tenenboim-Chekina, L. Rokach, Y. Elovici, Intruder or welcome friend: Inferring group membership in online social networks, vol. 7812, LNCS, Washington, DC, United states, 2013, pp. 368–376.
[30] J. Yang, J. Leskovec, Defining and evaluating network communities based on ground-truth, Beijing, China, 2012, pp. ACM Spec. Interest Group Knowl. Discov. Data (SIGKDD); ACM Special Interest Group on Management of Data (SIGMOD) –. .