Data Science and Engineering (2026) 11:84–99
https://doi.org/10.1007/s41019-025-00307-9

RESEARCH PAPERS

Opinion Maximization Based on Fairness in Social Networks

Yingying Zhai[1] · Zhenling Han[1] · Zefang Dong[1] · Xiaochun Yang[2] · Bin Wang[1,2]

Received: 22 January 2025 / Revised: 5 April 2025 / Accepted: 25 June 2025 / Published online: 17 November 2025

Abstract

Opinion maximization has attracted much attention in viral marketing. It selects an initial seed user set to disseminate user opinions on the target product and finally produces more positive opinions in social networks. In earlier studies, a critical but not studied problem is the fairness of information dissemination in groups with sensitive characteristic (such as age or race). People prefer to promote products for target users (majority groups) rather than in sensitive characteristic groups (minority groups). That leads to the differences in information dissemination between minority groups. In addition, social networks have the in-depth structural information. Therefore, in this paper, we design opinion maximization based on fairness framework (OMBF) using graph attention networks (GAT) to exploit more network information, and only consider the fairness in minority groups. OMBF composes of three parts: (1) the determination of candidate nodes according to node representations, (2) the dynamic changes in opinions, (3) the selection of final seed nodes. Firstly, we utilize GAT to obtain node representations and to determinate candidate nodes and design a node opinion formation model to model the dynamic changes in opinions. Then, we use the fair constraint value to ensure the fairness in the information dissemination process of minority groups. Based on above, final seed nodes are selected. We conduct experiments on synthetic and real-world datasets to show the effectiveness of our approach. The results indicate that the total opinions of active nodes in all nodes and fair values in minority groups are better than the chosen state-of-the-art benchmarks.

Keywords Social networks · Opinion maximization · Fairness · Node representations · Sensitive characteristic

1 Introduction

With the development of social system, more and more large social networks have become the most common social tools for crucial message propagation. Besides, social networks provide more and more new platforms for companies to promote their products in various scenarios. In recent years, influence maximization (IM) problem has obtained a great many of research interest in social network analysis, which is a key approach for viral marketing in modern business [1, 2]. The key of IM is mainly to find a set of initial influential nodes, also called seed users, using their influence to convince their social neighbors to adopt a target object, and aims at generating the maximum number of influenced users in social network [3–5].

In classical IM problem, for the target products, seed users and influenced users usually have positive opinions that are good comments. However, in real life, when seed users promote target products to their social neighbors, different neighbors would have different opinions. Opinion maximization (OM) is a practical extension of IM problem. It considers the difference of opinions for the target products, which are positive, negative and neutral. Therefore, OM aims to increase positive opinions in influenced users on target products and reduce the negative opinions, instead of maximizing the number of affected users. Accordingly, the characteristics and objective function of OM problem

✉ Yingying Zhai
zyy@mail.neu.edu.cn

Zhenling Han
zhenlingh@163.com

Zefang Dong
dongzefang@stumail.neu.edu.cn

Xiaochun Yang
yangxch@mail.neu.edu.cn

Bin Wang
wangbin@mail.neu.edu.cn

[1] College of Computer Science and Engineering, Northeastern University, Shenyang 110819, Liaoning, China

[2] School of Mathematics and Statistics, Liaoning University, Shenyang, China

have been changed relative to IM problem. In real life, OM could help more companies promote new products better. For example, a company pushes its new products to the market. It usually selects some potential customers to maximize the positive comments towards the products. By this way, the products can establish their good reputations, and more users can be attracted and gain the largest profit. Moreover, similar to IM, OM could be applied in many real world applications such as political votes [6], marketing management [7], advertisement dissemination in emerging networks such as VANETs [8, 9] and rumor blocking [10] in social networks.

Existing abundant studies on OM problem mainly focus on two aspects, seed nodes recognition and opinion propagation models. As for the recognition of seed nodes, the methods mainly include greedy algorithms [11–14], heuristic algorithms [15–19], as well as other variants [20–22]. The opinion propagation models, similar to IM, include independent cascade (IC) model [23], linear threshold (LT) model [24] and other new diffusion models improving feasibility of IC and LT [25]. However, the aforementioned studies face the following drawbacks: (1) they neglect the fairness of information dissemination in minority groups relative to other groups; (2) the consistency of fairness constraint condition for different groups would affect the goal of OM problem. In addition, as far as we know, there are no studies on OM problem taking fairness into consideration; and (3) only a few studies consider the structural information about social networks and user's attribute features. Also, people pay more attention to the active user's dynamic opinion. User's attribute features are crucial to most companies to promote new products. It largely determines the user's behaviors towards products. Thus, considering the users' attribute features is essential. Furthermore, for the inactive users, although they do not adopt the target product, they also would have some opinions on it. Thus, their opinions also affect the total opinions about the target products.

To address the aforementioned drawbacks, we propose opinion maximization based on fairness framework (OMBF) to enhance the fairness of information dissemination in minority. OMBF not only ensures the fairness of minority groups, that is, knowing the news of new products, but also maximizes the goal of OM problem. First, in this article, we utilize the GAT to mine deep network structural information and improve the accuracy of the determination candidate nodes. GAT learns the correlation in each pair nodes which could get node representations by attention mechanism, and the layer number of GAT and the number of multi-head determine how much network structural information that can be dug. Furthermore, we calculate the candidate nodes utilizing the obtained node representations. Second, we design a node opinion formation model considering the update of inactive nodes opinions. Finally, we set the fair constraint value in minority groups and on the premise of fairness to select the final seed nodes in all candidate nodes.

This paper summarizes the core contributions as follows.

(1) We formally define the OM problem, which is an optimization problem that considers fairness in minority groups. It is a NP-hard problem. What's more, it has no monotonicity and submodularity properties within a fairness constraint value. As we know, we are the first research using GAT to exploit network structural information to obtain candidate nodes in OM problem. Furthermore, we consider the fair constraint and inactive nodes opinions in social networks simultaneously.
(2) We design OMBF to solve OM problem, and it consists of three parts: (1) the determination of candidate nodes according to the node representations obtained by GAT, (2) the dynamic change in opinions including inactive nodes and (3) the selection of seed nodes considering a fair constraint value in minority. We leverage the similarity method to determine the candidate nodes according to node representations. Additionally, we set the fair constraint value to keep fairness in minority groups and select the final seed nodes.
(3) We prove our framework is effective via experiments in synthetic and real datasets. The experimental results show that our proposed approach has more total opinions and fair constraint values than other methods from the candidate set.

The rest of the sections are organized as below. Section 2 describes the related work about our study content. OM problem is formally defined in Sect. 3. Section 4 introduces the opinion maximization based on the fairness framework. Section 5 demonstrates experiment settings, and discusses relative results in detail. Finally, this paper in Sect. 6 concludes research.

2 Related Work

Related works are introduced in this section about three fields, including influence maximization problem, opinion maximization problem and graph representation learning.

2.1 Influence Maximization Problem

Influence maximization (IM) problem was studied at the earliest in [11], and Kempe et al. presented it as a discrete optimization problem. In addition, IM problem had been proven to be an NP-hard, and the objective function was monotonic and submodular. Furthermore, the authors designed an approximate optimal solution that was called

Greedy algorithm in two types of information propagation models, including IC and linear threshold (LT) in social networks. Since Greedy algorithm selects seed set from all inactive nodes in network in each iteration process, it has a high computation complexity issue and does not apply to large networks. Therefore, in order to decrease the run time on Greedy algorithms, some heuristic algorithms and their variants were proposed.

Leskovec et al. [3] considered submodularity property, and integrated cost-effective lazy forward optimization into framework that could decrease simulations number of Monte Carlo (MC), furthermore, reduce the time of node's spread in social networks. Simulation results demonstrated that improved algorithm CELF for IM problem was an efficient scalable solution, and it could decrease the complexity of Greedy algorithm. CELF++ [12], by reducing unnecessary MC simulation, reduced the amount of impact computation at the start part of CELF which divided the network and calculated the impact of different partitions. Chen et al. [4] designed a based on degree discount algorithm which could improve the impact propagation. Liu et al. [24] proposed GPR using page rank as the measure of influence spread to select seed nodes from candidate seed set. Despite their methods take the properties and information spread models, ignores the mutual impact of selected nodes. Borgs et al. [26] achieved a theoretical breakthrough and proposed a new algorithm in IC model, that is reverse influence sampling (RIS). The concept of using RIS, it considers a node often appears as an "influencer", and it could be a good candidate nodes for the final seed nodes.

In recent years, for solving IM problem, there are some using deep-learning and neural network approaches, which produce good results [27–29]. Such as, Keikha et al. [27] presented DeepIM algorithm using CARE [30] and skip-gram model [31] to obtain characteristic vectors about nodes and further to select seed nodes in embedding space. Zhang et al. [32] learned the node representations using the proposed fake labeling mechanism GCN method and further to determine seed nodes. He et al. [33] designed the two-stage iterative framework in IM problem which could eliminate nodes with less impact and reduce computational complexity. Besides, some works considered and investigated fairness. As described next. Fish et al. [34] studied individual fairness to minimize the access gap of spread information, however did not consider the group fairness constrain. Rahmattalabi et al. [35], studied influence maximization with group fairness for robust graph covering problems. Ali et al. [36] took consider time-critical into fairness in IM and addressed the IM problem through the proxy objective functions. Khajehnejad et al. [37] used adversarial graph embedding method to keep fairness in sensitive characteristic groups and achieved group fairness well. Khajehnejad et al. [38] proposed CrossWalk approach to enhance fairness in various graph algorithms, including the influence maximization problem.

From the research mentioned, we can see that in IM problem, the most active users default to consider positive opinions about the target products, and it aims to increase the amount of influenced users. In fact, different users have different opinions about the target products including positive, neutral, negative opinions. Thus, the OM problem arises to improve the IM problem. Existing fairness studies in IM, the fairness usually faces all groups, and the constraints of fairness are same to different groups, which would limit the goal of IM problem to some extent. Also, it is inconsistent with the fact that people pay more attention to the fairness in minorities.

2.2 Opinion Maximization Problem

In summary, OM problem aims to maximize total positive opinions towards the target products which promoted by companies in social networks. OM is a significant research subject and it has attracted widespread attention in social analysis. Besides, it integrates user's opinions into the traditional IM problem. A considerable amount of OM research is in social network. For example, Chen et al. [16] proposed a new information propagation model that extended the independent cascade model by considering negative opinions. This information propagation model devised a parameter to model negative behavior towards a product. For each user, the parameter was same, and it was impractical in most situations. People have different opinions towards the target product according to personal preference. In order to solve this issue, Zhang et al. [25] designed opinion-based cascading (OC) model according to the independent cascade model and further formulated OM problem on OC model which considered individual opinion value and captured the dynamic change in opinions simultaneously. In proposed OC model, the objective function had no submodular property, thus authors designed an available method to obtain more positive opinions.

Furthermore, Gronis et al. [18] designed the campaign problem to select seed user set which had positive opinions about target products and maximized positive opinions in social network. People designed an efficiency approach to solve the OM problem. Abebe et al. [19] researched overall opinions in budget and no budget constraints. For the first, it proved that NP-hard, no submodular and no supermodular properties. It utilized the proposed approach, which could obtain the more marginal gain to solve the problem. For the second, it showed that in a polynomial-time approach, OM could obtain the results. The active opinion maximization (ATOM) was studied by Liu et al. [14], and the goal of it was to select some nodes and to produce the opinions in multi-round spread. It based on user preference data to know user opinions which was different from previous works that consumed the user opinions were known. The authors designed CONE framework considering opinion estimation and influence propagation to settle the ATOM problem. In a recent work, He et al. [39] designed the multi-stage heuristic algorithm (MHA) to obtain seed users. Further, He et al. [40] formalized the competitive opinion maximization (COM) problem in signed social networks. In order to address competitive problem, the authors designed the reinforcement-learning algorithm to select the seed nodes. He et al. [41] proposed QDOM to consider dynamic OM problem. QDOM includes two parts, first is opinion model which could obtain the dynamic changes in opinions using stateless Q-learning theory, and second is Q-learning based seeding algorithm to select seed nodes.

In summary, there are a few works that study the fairness of information dissemination in minority groups, most works consider the same fairness constraints in all groups, which is inconsistent with the reality that minority groups are more concerned. Further, the existing OM-based approaches do not consider the fairness in social networks. Besides, most OM algorithms based on heuristic approaches. The utilization of network structure information determined the accuracy and stability of the heuristic algorithm. The less network structure information, the accuracy of heuristic algorithm cannot be guaranteed. Inspired by the above limitations, we utilize GAT to obtain the network structural information, it can greatly improve the efficiency of selecting final seed nodes. And only consider fairness of information dissemination in minority groups that ensures the goal of OM to a certain extent.

2.3 Graph Representation Learning

Network graph structure is a nonlinear data structure and massively presents in real life. There are existing research areas, such as transportation network, subway network, social network, state execution (automaton) in computer, etc., which can be abstracted into graph structure. Obtaining graph nodes low-dimensional representation in social networks is the first part of our proposed framework.

Graph representation learning aims to convert nodes to lower dimensional space which could reconstruct network topology information, could preserve node's topological information, and could be applied to common applications in social networks. There are most approaches, which could be classified three classes, including based on random-walks [42, 43], deep learning architectures [44, 45], and graph neural networks [46]. In this article, the method most related to our work is GAT [47], which is an approach of graph neural networks for deriving node representations and could obtain more network structural information.

GAT applies the attention mechanism to graph representation learning in message spread process and using a self attention strategy to compute each node representation, which could obtain better node embeddings. Therefore, we utilize GAT to the social relationship to get better node representation in social graph network.

GAT based on graph convolutional network (GCN) [47] and incorporated self-attention mechanism that leverages attention coefficients in self attention layers for each node. For the nodes self-attention mechanism is a shared mechanism. atten: $\mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ and F' donates characteristic dimension in each node, and the attention coefficients between each pair of nodes are computed as:

$$e_{ij} = atten(Wh_i, Wh_j) \tag{1}$$

where $h_i, h_j \in \mathbb{R}^F$ represents nodes features and $W \in \mathbb{R}^{F'} \times \mathbb{R}^F$ is a weight matrix. atten usually is a dot product and a non-linearity function. In order to easily calculate coefficients and compare across different nodes, usually the softmax function is used to normalize the attention coefficients:

$$a_{ij} = softmax_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \tag{2}$$

where N_i denotes the neighbors of node i in Graph. e_{ij} donates the impact intensity of node j on node i. By incorporating neighborhood information and a_{ij}, applying a non-linearity function σ, the output feature h_i' can be formed as:

$$h_i' = \sigma\left(\sum_{j \in N_i} a_{ij} W h_j\right) \tag{3}$$

Multi-head attention can make the learning process of self-attention more stable, thus, it is necessary to use multi-head attention to replace the attention mechanism. K independent attention mechanisms perform the variation of the Eq. 3, and then concatenating the features, the final output feature represents as follows:

$$h_i' = \|_{k=1}^{K} \sigma\left(\sum_{j \in N_i} a_{ij}^k W^k h_j\right) \tag{4}$$

where a_{ij}' are the normalized attention coefficients of kth attention mechanism. In this paper, we utilize GAT to generate node representations, and through the training of GAT based on supervised learning, which could obtain better node representations. Better node representation can be used to compute similarity of nodes in embedding space, which could obtain better candidate nodes.

3 Model Definition

3.1 Problem Model

This paper formally defines the opinion maximization based on fairness in minority group: A directed connected graph $G = (V, E, W)$ represents the social network, and $V = \{v_1, v_2, \dots, v_N\}$ (N is the number of nodes) denotes nodes set in social networks, and $E = \{e_1, e_2, \dots, e_M\}$ (M is the number of edges) is the edges set. In graph G, users are represented by nodes and users' social associations are represented by edges between nodes. $W = \{w_{ij}, i, j \in [1, N]\}$ denotes the strength of association and reveals the significance from v_i to v_j.

In most cases, IM problem selects k seed nodes to disseminate information and finally to obtain more influenced users in social networks. Different from it, the OM aims to maximize the positive opinion of influenced nodes in the process of information dissemination. Furthermore, only several works of IM consider the fairness, which mostly is the same constraints for different groups that is impractical in real life, where people care more about the requirements in minority groups. As we know, the OM problem had not research the fairness in information dissemination in social networks. Therefore, we are the first to study the OM problem based on fairness (OMBF) and in order to describe the OMBF more specific, in the following notations are given.

Definition 1 (*Seed node*) According to the initial budget condition, select k nodes as the initial seed nodes with a certain influence could propagate opinion about the target product to their neighbors via the social relationships in graph G. Seed nodes are the optimal set that can maximize the interests of businesses, thus, selecting seed nodes is an important issue in social networks.

Definition 2 (*Fair constraint value*) The fair constraint value is a fair guarantee in minority groups. In this paper, we define it as the change in opinion after information propagation relative to the initial opinions in minority groups, which means the goal of fair is to let users know the information about the promoted target product and can form certain opinions. It can prevent the spread of false information to a certain extent because it takes into account the fairness of information dissemination in minority groups. Also, we use the change of opinion to measure the degree of knowing about product information in minority groups.

Definition 3 (*Node opinions*) All nodes have initial opinions that vary between $(-1, 1)$, where ranging $(0, 1)$ is a positive opinion and 0 denotes the neutral opinion, others are negative opinion. The node i opinion is shown as o_i. Each active nodes has the change to influence neighbors around it from inactive to active status. When inactive neighbors are influenced by active nodes, the neighbor's opinion always change even though neighbors are inactive. That is to say, the active nodes propagate opinions to inactive nodes which opinions would change in two ways depending on the status of the nodes (active or inactive). We respectively leverage V_1^+ and V_a^+ to stand for the set of all nodes that belong to minority groups and active nodes in graph G with positive opinions towards target products, V_1^0 and V_a^0 to denote the set of all nodes in minority groups and active nodes in graph G with neutral opinions, and the V_1^- and V_a^- to denote the set of all nodes that belong to minority groups and active nodes in graph G with negative opinions.

Definition 4 (*Fairness metric*) Fairness ensures that minority groups (V_1) receive sufficient opinion enhancement during information propagation. We define fairness as:

$$\text{Fairness} = \frac{\sum_{v_t \in V_1} (o_t - s_t)}{\sum_{v_t \in V_1} |s_t|} \tag{5}$$

Here, the numerator represents the total positive opinion change (difference between post-propagation opinion o_t and initial opinion s_t), while the denominator is the sum of absolute initial opinions. The constraint Fairness $\geq \beta$ guarantees that the proportional positive enhancement in minority groups' opinions meets a threshold β. This metric avoids extreme opinion flips (e.g., $+1$ to -1) and focuses on meaningful improvements in minority opinions.

Definition 5 (*Opinion maximization*) Given graph G, we use the initial fair constraint value to select k seed nodes that ensure the fairness of information dissemination in minority groups. Using the node opinion formation model, seed nodes set S propagates their opinions about the target product to their inactive neighbors and aims to produce more opinions of active nodes. Accordingly, our formal definition of OM problem is as follows:

– maximize:

$$F(S) = \sum_{v_i \in V_a^+} o_i + \sum_{v_j \in V_a^-} o_j \tag{6}$$

– subject to:

$$\frac{\sum_{v_t \in V_1} (o_t - s_t)}{\sum_{v_t \in V_1} |s_t|} \geq \beta \tag{7}$$

where $F(S)$ is the objective function and it aims to maximize positive opinions in all active nodes which implies have a good reputation for the target product. The numerator $\sum_{v_t \in V_1} (o_t - s_t)$ represents the total change in opinion values of minority group nodes before and after information propagation, while the denominator $\sum_{v_t \in V_1} s_t$ is the total initial opinion values of minority group nodes (including both positive and negative). In this fairness constraint, we encourage positive shifts in the opinions of minority group nodes. We refer to the ratio of this total change to the total initial opinions as the fairness value of the social network. And the β is the fair constraint value to maintain the fairness of information dissemination in minority groups.

3.2 Opinion Formation Model

3.2.1 Active Nodes

The opinion linear threshold (OLT) model extends traditional LT and IC models by integrating opinion dynamics [23, 24]. Nodes within the network are categorized into two states: active or inactive. A node becomes active under two primary mechanisms. First, through seed selection, where nodes initially chosen as part of the seed set S trigger the propagation process. Second, via neighbor activation: an inactive node v transitions to active at time t if the aggregated influence from its active in-neighbors surpasses its activation threshold θ_v. Formally, this occurs when $\sum_{u \in A_{t-1}} w_{uv} \geq \theta_v$, where A_{t-1} denotes the set of nodes active at time $t-1$, and w_{uv} represents the edge weight from u to v. Once activated, a node remains permanently active and propagates opinions to its neighbors in subsequent time steps.

Inactive nodes retain their state until activated but may adjust their expressed opinions through interactions with active neighbors, as detailed in Sect. 3.2.2. The state transition dynamics emphasize irreversibility: activation is a one-way process, and nodes cannot revert to an inactive state. Both seed nodes (initially active) and nodes activated through neighbor influence collectively drive the propagation process, ensuring continuous expansion of the active set over time.

3.2.2 Opinion Change

We model the opinion dynamics of nodes during information dissemination, accounting for both active and inactive influenced nodes. Each node v maintains a fixed intrinsic opinion s_v, which remains unchanged throughout the process. However, its expressed opinion o_v adapts based on its activation state after exposure to active neighbors. When an inactive node v becomes activated, its opinion updates as a weighted average of its intrinsic opinion and the influence from active in-neighbors:

$$o_v = s_v + \frac{\sum_{u \in N_v^{in}} o_u w_{uv}}{|N_v^{in}|} \tag{8}$$

where N_v^{in} is the set of in-neighbors of v, o_u is the opinion of active neighbor u, and w_{uv} quantifies the influence strength from u to v. The term $\frac{1}{|N_v^{in}|}$ ensures normalization over all influencers. If v remains inactive despite influence attempts, its opinion adjusts proportionally to the impact of the most recent active neighbor b, scaled by the total absolute influence potential of all in-neighbors:

$$o_v = s_v + \frac{o_b w_{bv}}{\sum_{u \in N_v^{in}} |o_u| w_{uv}} \tag{9}$$

Here, o_b is the opinion of the active node b, and the denominator normalizes the influence by the cumulative weighted opinions of all in-neighbors, ensuring the adjustment reflects relative influence strengths.

This dual-state mechanism ensures that active nodes propagate opinions, while inactive nodes still update opinions, thereby preserving fairness by representing minority groups even if they resist activation. To model such resistance, minority group nodes are assigned higher activation thresholds ($\theta_v \in (0.5, 1)$), which delay their activation but still allow opinion shifts through Eq. (8). Compared to traditional propagation models, this approach better aligns with real-world scenarios where users dynamically adjust their attitudes based on social influence. It enables promoters to track real-time reputation shifts and strategize interventions effectively.

4 Opinion Maximization Based on Fairness Framework

Figure 1 illustrates the process of selecting final seed nodes in proposed OMBF framework. Firstly, according to network structure information and node attributes, we use GAT to obtain all node representations and utilize similarity method to determine candidate nodes in embedding space. Next, use the fair constraint value to obtain the final seed nodes and begin information propagation. Algorithm 4 summarizes the whole process of OMBF. Specifically, to describe the detail of OMBF, we first present how to determine the candidate nodes, and then we design the node opinion formation model to capture dynamic changes in opinions. According to the candidate nodes and the fair constraint value, we select the final seed nodes to propagate information.

Firstly, we present the classical greedy algorithm in OM problem in Algorithm 1. The algorithm utilizes the following marginal gain to determine in all remaining nodes which node v could be added to set S. When the size of the set S

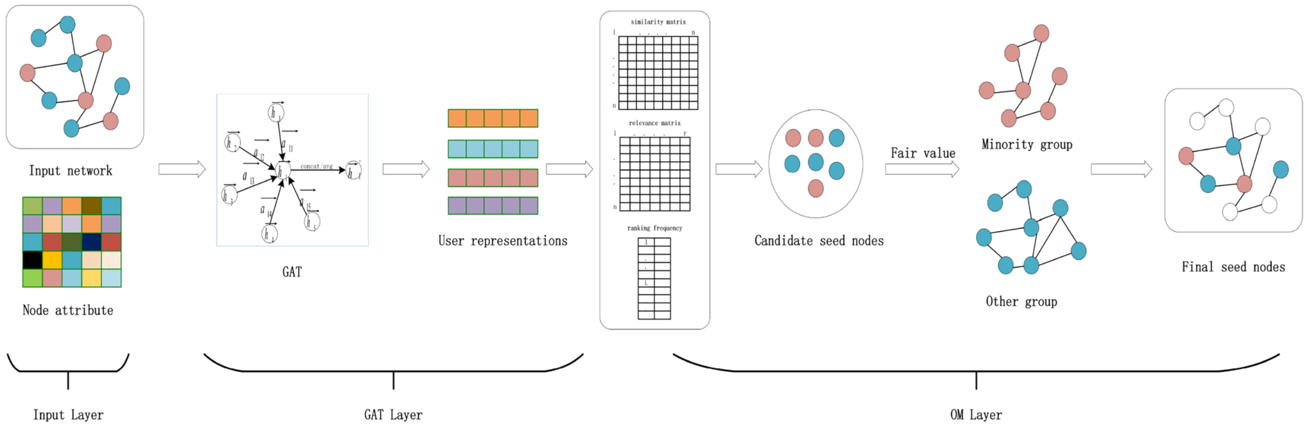


Fig. 1 Opinion maximization based on fairness framework

reaches k or no more nodes could be activated, the algorithm would terminate and return the final seed nodes.

$$\varepsilon(S) = T(S \cup \{v\}) - T(S) \tag{10}$$

where $T(\cdot)$ denotes the sum of opinions of all active nodes when seed nodes are given. From above, we notice the greedy algorithm calculates the optimal gain in all remaining nodes, which could result the efficiency that consumes a long time to select nodes. To solve the above problem and obtain superior opinion value, we design the OMBF framework to solve the OM problem in our work. Next, we will continue describe the process of OMBF.

Algorithm 1 Greedy algorithm

Input: Graph $G = (V, E, W)$, and seed budget size k
Output: Seed node set S
Initialize set $S = \varnothing$;
for $i = 1$ to k **do**
  for $v \in V - S$ **do**
    Calculate the optimal gain v' by Equation 10;
  end for
  $S \leftarrow S \cup \{v'\}$;
end for
return S

4.1 Determination of Candidate Nodes

Considering existing graph network data contains rich information and the properties of network nodes could reflect the user behavior in some degree. Extracting more network structural information could solve OM task better. Therefore, we first leverage the GAT model to exploit the more network structural information and obtain node representations. The GAT model uses attention coefficients to assign weights sum and aggregates neighbors' information for each node, so we can extract more information from the network by setting more GAT layers and multi-head. Second, we calculate the similarity between nodes in embedding space according to node representations obtained by the GAT model, which is similar to the method in the work [27]. However, the work employed deep learning techniques to extract node representations in a network, which is different from us. We utilize the cosine similarity to compute relatedness to identity relevant nodes of each node. According to node representations, the similarity between node u and node v could be computed as follows:

$$conse(u, v) = \frac{f(u) \cdot f(v)}{|f(u)||f(v)|} = \frac{\sum_{i=1}^{d} f(u)_i f(v)_i}{\sqrt{\sum_{i=1}^{d} f(u)_i^2}\sqrt{\sum_{i=1}^{d} f(v)_i^2}} \tag{11}$$

where $f(u)$ and $f(v)$ are node u, v representations. We calculate relatedness between each node and others. After that, we select r nodes which have high relatedness for each node in the graph network, and then, from $N \times r$ nodes with high relatedness, we rank L nodes with the most occurrences as candidate nodes.

In summary, in this paper, we employ Algorithm 2 to identify candidate nodes by leveraging the expressive power of graph attention networks (GATs). The algorithm operates in three key phases, each contributing to the robust selection of influential nodes. First, the GAT model generates node representations by aggregating neighborhood information through multi-head attention mechanisms. This phase captures both local and global structural patterns, with each attention head computing weighted coefficients as follows:

$$\alpha_{vu}^{k} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\mathbf{a}_k^T[\mathbf{W}_k f(v) \| \mathbf{W}_k f(u)]\right)\right)}{\sum_{w \in \mathcal{N}(v)} \exp\left(\mathrm{LeakyReLU}\left(\mathbf{a}_k^T[\mathbf{W}_k f(v) \| \mathbf{W}_k f(w)]\right)\right)} \tag{12}$$

where k indexes the attention heads. The final representation $f'(v)$ is derived by concatenating or averaging the outputs from all heads, enabling multi-scale feature extraction. In the second phase, the algorithm quantifies node relatedness in the embedding space using the cosine similarity (Eq. 10). For each node v, this involves computing pairwise similarities with all other nodes and selecting the top-r most related nodes. This step ensures that the candidate pool reflects both direct and indirect influence relationships. The third phase ranks the $N \times r$ high-relatedness nodes by their occurrence frequency across all top-r lists. By selecting the top-L nodes with the highest occurrences, the algorithm prioritizes nodes that consistently exhibit strong connections to diverse regions of the network, thereby maximizing their potential as influential seeds. This multi-layered approach combines GAT-driven representation learning with graph-theoretic centrality measures, offering a principled balance between local relevance and global prominence.

Algorithm 2 Selection of candidate nodes

```
Input: G = (V, E, W), the high relatedness number of each node r and the number
    of candidate seed nodes L
Output: Candidate nodes set
Obtain node representations using GAT model;
for v ∈ V do
    for u ∈ V − v do
        Calculate the similarity between v and u by Equation 11;
    end for
    Rank r nodes most related to v;
end for
Calculate the occurrences of the most related N ∗ r nodes;
Select the top L nodes according to the occurrence in related nodes as candidate
seed nodes;
return Candidate seed nodes
```

4.2 Node Opinion Formation Process

Given the graph G, after the determination of candidate nodes in embedding space, the final seed nodes are further determined and will propagate the opinions towards the target product in social networks. Thus, designing an effective approach to estimate the node's opinion is crucial. Previous approaches considered the opinion change in active nodes and opinion formation process is static. Different from them, in this paper, we take into account the changes in the opinions of inactive nodes and consider dynamic opinions in the information propagation process. According to the information propagation model OLT, Algorithm 3 demonstrates the pseudo-code of the node opinions change process, including three parts: (1) initialize the active nodes set (line 1); (2) according the information propagation model (OLT) to judge nodes state (lines 2–4); (3) update the opinion of two state nodes utilizing node opinion formation model (lines 5–11).

Algorithm 3 Node opinion formation process

```
Input: G = (V, E, W), seed nodes set S
Output: The updated opinions of influenced nodes
 1: Initialize set A = S, S' = ∅;
 2: while S ≠ NULL do
 3:     for v ∈ V − A do
 4:         if ∑_{u∈S} w_uv > θ_v then
 5:             Update the opinion of node v by Equation 8;
 6:             A ← A ∪ {v};
 7:             S' ← S' ∪ {v};
 8:         else
 9:             Update the opinion of node v by Equation 9;
10:         end if
11:     end for
12:     S ← S';
13:     S' ← ∅;
14: end while
15: return The updated opinions of influenced nodes
```

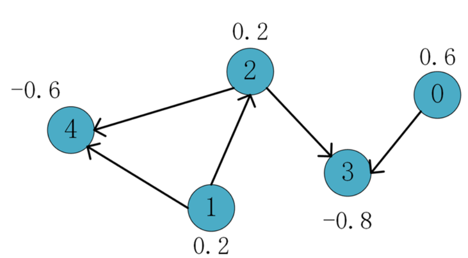


Fig. 2 An example of non-monotonicity and non-submodularity

4.3 Selection of Final Seed Nodes

Through obtaining candidate nodes, we could select the final seed nodes using the fair constraint value. We guarantee the fairness of information dissemination in the minority groups by the fair constraint value. Algorithm 4 summarizes the whole process of OMBF, which consists of five parts: (1) we initialize seed nodes set (line 1); (2) we calculate the L candidate nodes (line 2); (3) we rank candidate nodes L_1 that belong to minority groups using the gain of opinion absolute value (line 3); (4) we rank candidate seed nodes L_2 that belong to other groups using the gain of opinion (line 4); (5) for the minority group, we select node one by one in the minority group candidate nodes into the seed nodes set s_1, and remaining nodes selecting in the other group candidate nodes into the seed nodes set s_2, stopping until the sensitive group satisfies the initial fair constraint value. (6) the final seed nodes $S = s_1 \cup s_2$ (lines 5–13).

The time complexity of Algorithm 4 (OMBF) is primarily determined by two components: candidate seed selection (Algorithm 2) and the seed screening process (steps 5–15). Algorithm 2 employs a GAT model to compute node representations and select top-L candidate nodes, with a time complexity of $O(|V|^2 \cdot d + |V| \cdot r \log r + |V| \cdot L)$, where d denotes the embedding dimension of GAT and r represents the node relevance calculation range. The seed screening phase requires iterating through the minority group candidate set L_1 (of size $O(L)$), with each iteration involving fairness constraint validation (Eq. 6) and seed set merging, yielding a time complexity of $O(L \cdot k \cdot T_{\text{fair}})$, where T_{fair} denotes the computation time for fairness verification. Consequently, the overall complexity of the algorithm is $O(|V|^2 \cdot d + L \cdot k \cdot T_{\text{fair}})$, dominated by node representation computation and iterative fairness constraint validation.

Algorithm 4 Opinion maximization based on fairness (OMBF)

Input: Network node set $G = (V, E, W)$, the number of seed nodes k, L candidate seed nodes and fair constraint value β
Output: Seed node set S
Initialize set $S = \varnothing$, $s_1 = \varnothing$, $s_2 = \varnothing$;
Calculate L candidate seed nodes by Algorithm 2, $L = L_1 \cup L_2$, where L_1 belongs to minority group, L_2 belongs to other group;
Rank L_1 in descending order by Equation 14 in minority group;
Rank L_2 in descending order by Equation 15 in G;
for $i = 1$ to $|L_1|$ **do**
  $s_1 \leftarrow s_1 \cup \{i\}$;
  $L_2[: (k - |s_1|)]$;
  $S = s_1 \cup s_2$;
  if Equation 7 **then**
    Break;
  end if
  if $s_1 = |L_1|$ **then**
    Break;
  end if
end for
return Seed node set S

Here, the proposed gain of opinion absolute value is to ensure minority groups more quickly satisfy the initial fair constraint value. First, the absolute opinion value is defined as:

$$T = \sum_{v_i \in V_1^+} \left|\Delta o_i\right| + \sum_{v_j \in V_1^-} \left|\Delta o_j\right| \tag{13}$$

where T denotes the after information propagation total change of absolute opinion value. After that, the gain of opinion absolute value could denote as following, which is similar to the core of Greedy algorithm:

$$\varepsilon(s_1) = T\left(s_1 \cup \{v\}\right) - T(s_1) \tag{14}$$

Next, we introduce the gain of opinion, which is similar to the above definitions and could be defined as:

$$\varepsilon(s_2) = F\left(s_2 \cup \{v\}\right) - F(s_2) \tag{15}$$

and $F(\cdot)$ is the objective function.

In the end, we subsequently introduce two Theorems that are related to us, as shown below.

Theorem 1 *The OM problem in definition 4 is NP-hard.*

Proof According to the NP-hardness theory [25], when a problem in a restricted class instance is NP-hard, then it is also NP-hard. Since the traditional IM problem is NP-hard [11], we consider a restricted instance under our opinion formation process. For each node u, we set o_u to 1, and the fair constraint value β is set to 0. Then when u is activated by its neighbors v, the node u will not change opinion, it will hold the same opinion. Without opinion change, therefore, maximizing the active nodes opinions equals to the classical IM problem. So, in our paper the opinion maximization based on fairness is still a NP-hard problem. □

Theorem 2 *The objective function $F(\cdot)$ in this paper is non-monotone and non-submodular.*

Proof The monotonicity needs to function f satisfy the inequality $f(T) \geq f(S)$ for all $S \subseteq T$. Because of considering the negative opinions in user opinion, when activating nodes with negative opinions which will decrease the total opinions and does not satisfy monotonicity. The function f is submodular if $f(S+v) - f(S) \geq f(T+v) - f(T)$ for all $S \subseteq T \subseteq V$, $v \in V$ and $v \notin T$. We give counter examples for non-monotonicity and non-submodularity respectively to prove Theorem 2. As illustrated in Fig. 2. The innate opinion of node0–node4 is $0.6, 0.2, 0.2, -0.8, -0.6$ respectively, the influence threshold is 0.1 for each node, and the weight is equality to 1/in-degree. Here, we according the Eq. 8 to calculate the opinion of active nodes. □

Non-monotonicity. Suppose that $S_1 = 1$, $S_2 = 1, 3$, $S_3 = 0, 1, 3$, then $F(S_1) = 0.1$, $F(S_2) = -0.7$, $F(S_3) = -0.1$. As $F(S_1) > F(S_3) > F(S_2)$, so $F(\cdot)$ is non-monotone.

Non-submodularity. Suppose that $S = 1$, $T = 1, 3$, then $F(S+0) - F(S) = 0.05$, and $F(T+0) - F(T) = 0.6$. Because $S \subseteq T$ and $F(T+0) - F(T) > F(S+0) - F(S)$, $F(\cdot)$ is non-submodular.

5 Experimental Results

The experiment is conducted on both synthetic and real datasets with our propose framework and benchmark approaches to evaluate the performance. Firstly, we describe the datasets and experimental setup. Next we describe the baseline comparison method. Finally, we demonstrate results analysis.

5.1 Data and Experiment Setup

We utilize real and synthetic social networks and Table 1 describes them, including:

- *Rice-Facebook* The real dataset indicates students social relationship in Rice University, collected by Mislove et al. [48]. The student social graph contains students

Table 1 Datasets

Network	Nodes	Edges	Average degree
Rice-Facebook	1205	42,443	70.4
Antelope-Valley	500	1689	3.4
Cora	2708	5429	3.9
Twitter	3753	13,986	7.45

information such as age (which is a number between 18 and 20), college as well as major, and it is an undirected graph. We regard student's age as sensitive characteristic and consider students with ages 18 or 19 as group A that is the minority group, and with ages 20 as group B, which excluding node age higher than 20. In this paper we double the edges from undirected to directed graph.

- *Antelope-Valley* The synthetic dataset is one of the network used by Wilder et al. [49] to model an obesity prevention intervention in the Antelope Valley region of California. The dataset is a directed graph, and in this network each node owns geographic location, age, race, gender, as well as weight status. Considering the public health problem, obesity is an crucial factor, thus we select the weight status as the sensitive characteristic. The network has 500 nodes and we regard the weight status is normal as group A that is the minority group, and others as group B.
- *Cora* The real dataset is an undirected graph composed of machine learning papers [50], which nodes denote papers and edges represent reference relationships between papers. This dataset contains seven categories papers, and has 1433 dimension features. We consider the Neutral Networks as the sensitive characteristic, which is one of the categories in graph. We also double edges and make it become a directed graph.
- *Twitter* [51] The dataset, a subset of Twitter network comprising approximately 3753 nodes and 13,986 directed edges, identifies "neutral" political leaning as the sensitive characteristic (minority group) for analyzing fairness in information dissemination.

In our work, we only consider two groups on one sensitive characteristic. For graphs we tested on, we assign the weight according to the number of neighbors, where for every incoming edge of node v defined $1/N_v^{in}$ as the weight. We use the random method to generate initial user opinions in the range $(-1, 1)$. Moreover, for the user's threshold, in the minority groups, the threshold is between (0.5, 1), others is range (0, 1). Minority groups have a high threshold in inactive nodes, however, this does not affect the change of their opinions towards the target product. That is, when nodes in minority groups are promoted by other active nodes, they also engender some opinions on the product, although they are an inactive state. Due to user opinions being randomly generated, we run an average of 10 times on algorithms on each network.

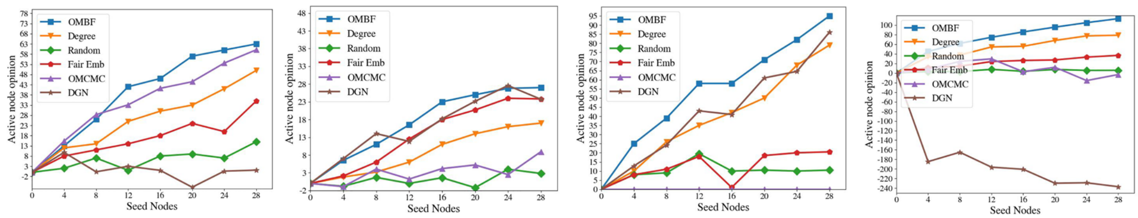
For the four datasets used in this study, minority and majority groups are partitioned based on sensitive characteristics, following real-world fairness studies [52–54]: (1) Antelope-Valley: users with normal weight status (majority) are distinguished from overweight/obese users (minority). (2) Rice-Facebook: users aged ≤ 19 (minority) are separated from those aged 20 (majority). (3) Cora: Papers in the "Neural Networks" category (minority) are differentiated from others (majority). (4) Twitter: Users with neutral political views (minority) are labeled separately from those with partisan leanings (majority).

In order to train the GAT model [47] to obtain node representations, we mark Group A as 0 and Group B as 1. Besides, we set the parameter as follows. The dimension of the output layer is set to 2, denoting groups class. The dimensions of the input layer are consistent with the number of node attributes, node threshold and node opinion. For the hidden layers, the Rice-Facebook is set to 3, the Antelope-Valley is set to 4, the Cora and Twitter are set to 8. The learning rate is set to 0.05, and dropout with 0.6 is applied. In Algorithm 2, r represents the number of nodes with high similarity to each node. During the simulation, r is set to 150 for the first two datasets and 510 for the latter two datasets. Additionally, L determines the number of candidate nodes: it is set to 90 for the first two datasets and 900 for the latter two datasets. To ensure fairness in information dissemination for minority groups, β is fixed at 0.001 for the first two datasets and increased to 0.01 for the latter two datasets. This adjustment aims to promote accurate information awareness in minorities, where opinion dynamics are used to measure their exposure to information.

5.2 Baseline Algorithms

Our proposed algorithm OMBF is compared with the following five base algorithms (i.e., Degree [55], Random [56], Fair Emb [37], OMCMC [57], DGN [58]). Specifically, the algorithms information is introduced as follows:

- *Degree* [55] Degree demonstrates the amount of first-order neighbors. The higher the degree a node has, the more it implies a great influence power.
- *Random* [56] Randomly select k nodes as seed nodes to propagate information about the target product in the graph.
- *Fair Emb* [37] Leveraging auto-encoder to embed graph information, and a discriminator to identify different



(a) Rice-Facebook (b) Antelope-Valley (c) Cora (d) Twitter

Fig. 3 The active nodes opinions of different approaches over the seed nodes on four networks

groups, which produces similarity distributed across sensitive characteristics. Then, to select seed nodes set by clustering in the embedding space.

- *OMCMC* [57] Leveraging centrality metrics to rank node importance and clustering to group structurally similar nodes, which produces candidate seed sets with balanced local and global influence. Then, to refine the final seed selection using objective functions based on network properties.
- *DGN* [58] Leveraging dual coupled graph neural networks to embed structural and attribute information, and a deep Q-network to learn adaptive influence-aware policies, which produces dynamic node representations optimized for long-term influence rewards. Then, to iteratively select seed nodes via reinforcement learning without predefined heuristics.

In addition, so as to better denote the performance of OMBF, we choose the classic Greedy algorithm as the baseline for the experimental part. Furthermore, we compare the efficiency of different algorithms in the same opinion value and information propagation mode OTL with OMBF.

5.3 Result Analysis

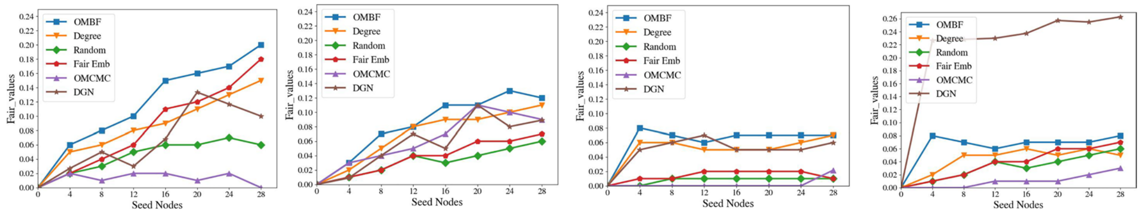
We measure OMBF performance with five baseline algorithms across four datasets and compare their total opinions in active nodes, fair values in group A, along with the comparison of global opinion expression values in the network after propagation. Then, we compare OMBF with the prevailing Greedy algorithm and show its effectiveness. At last, we set the same initial opinions and compare OMBF with other approaches to solve the randomness of initial opinion setting and further prove the effectiveness of OMBF which could maximize the positive opinions after propagation on the premise of fairness in minority groups with sensitive characteristics.

5.3.1 Active Nodes Opinions Over the Number of Seed Nodes

Firstly, we compare the sum of opinions in active nodes influenced by seed nodes during the information dissemination process. Figure 3 shows the spread of opinions of selecting seed nodes from 0 to 28 in three networks. We can notice that the total opinions of different methods are steadily growing as the number of seed nodes increases. Particularly, OMBF has larger total opinions than the other five baseline algorithms almost in each seed budget on real and synthetic datasets. Specifically, for the random approach obtains the smallest opinions because of the irregular selection of seed nodes. Besides, considering the negative opinions, the objective function is not monotonically increasing. Therefore, the total opinions do not necessarily increase over each seed nodes budget. For example, on the Antelope-Valley network, total opinions (Fair Emb, Random) drop when seed nodes varies from 20 to 24. In general, from Fig. 3, it observers the budget size of seed node could impact total opinions, and the OMBF method could select initial seed nodes effectively and obtain larger opinions than other methods.

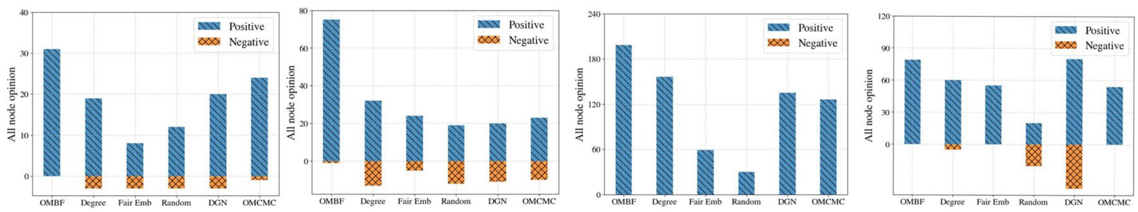
5.3.2 Fair Values in Group A Over the Number of Seed Nodes

Figure 4 illustrates the fairness values of different methods under varying seed node budgets. Results show that fairness values of all methods generally increase with more seed nodes. In most cases, OMBF outperforms other methods across nearly all seed node budgets on the four networks. Notably, on the Twitter dataset, OMBF exhibits slightly lower fairness values compared to DGN. However, DGN's total active opinion value is significantly lower than OMBF and all baseline algorithms, indicating its inefficiency in improving the information reach for minority groups. Fairness values reflect the proportion of opinion shifts relative to initial opinions, aiming to ensure minority groups can access and form specific opinions toward the target product. It is worth noting that fairness values are not strictly monotonic. For instance, on the Antelope-Valley network,



(a) Rice-Facebook (b) Antelope-Valley (c) Cora (d) Twitter

Fig. 4 The fair value in group A of different approaches over the seed nodes on four networks



(a) Rice-Facebook (b) Antelope-Valley (c) Cora (d) Twitter

Fig. 5 The groups opinions of different approaches after propagation over $k = 28$ on four networks

the fairness values of Fair Emb and Random drop when seed nodes increase from 12 to 16. Two factors explain this: first, the initial opinions of nodes are randomly generated; second, the stochastic propagation of opinions during information dissemination may lead to non-strictly increasing fairness values. Overall, we can clearly find that OMBF could acquire effective seed nodes and keep high fairness in the minority groups.

5.3.3 Comparison of Global Opinion Expression Values in Network

Figure 5 compares the global opinion expression values in the network (based on the same initial opinion) of OMBF and five other algorithms under a seed budget of 28. Experiments show that OMBF significantly outperforms the comparative algorithms: It achieves the highest positive opinion value and the lowest negative opinion value in the network after propagation, demonstrating the optimal overall performance. Among them, Degree, DGN, and OMCMC show close performance, while Fair Emb and Random yield poorer results. On the Twitter dataset, although the positive opinion values of OMBF and DGN are similar, DGN's negative opinion value is significantly higher, highlighting OMBF's advantage in balancing positive and negative influences. To conclude, the results imply the effectiveness of OMBF in selecting seed nodes that can make different groups gain the largest opinions after the propagation and facilitate the promotion of the target products.

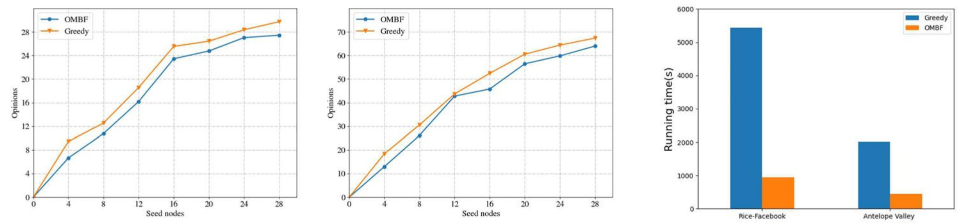
5.3.4 Comparison with Greedy Algorithm

In order to better display the performance of our OMBF approach, we compare it with Greedy algorithm which is generally superior to heuristic algorithms on the efficiency. In our experiment, we also use the same propagation model after selecting seed nodes with Greedy algorithm. Figure 6a, b demonstrate the total opinions of OMBF and Greedy on real and synthetic network. The results indicate Greedy algorithm shows slightly more opinions than OMBF with the increase in seed nodes and the total opinions in active nodes are very close to OMBF. However, there is a significant difference between running time in two algorithms in Fig. 6c. OMBF has considerably lower running time than Greedy algorithm. Here, we set seed nodes budget to 28, and run once in Rice-Facebook dataset and ten times in Antelope-Valley dataset, which could in one figure to better show and see the running time. In summary, Greedy algorithm selects seeds from the remaining nodes according to the objective function each time, so that the objective function can obtain a better value, but it spends much running time. Different form that, the OMBF algorithm determines candidate nodes using node representations by GAT model in embedding space in advance, reducing the computational complexity, and obtaining the total opinions.

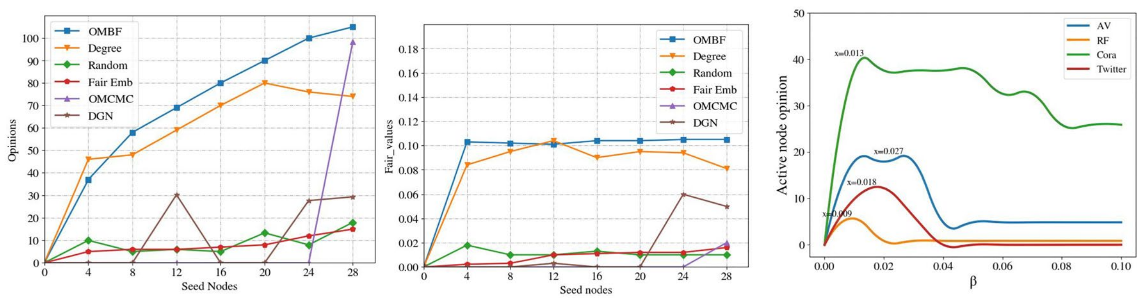
5.3.5 Comparisons in Same Initial Opinions

To further demonstrate the performance of OMBF, we test under the same conditions, that is, set the same initial

Fig. 6 Performance comparison of OMBF and Greedy: **a** active node opinions on the Rice-Facebook dataset, **b** active node opinions on the Antelope-Valley dataset, **c** runtime comparison between OMBF and Greedy



(a) Rice-Facebook (b) Antelope-Valley (c) Runtime comparison



(a) Active nodes opinions (b) Fair value (c) Parameter Analysis

Fig. 7 **a**, **b** Nodes with the same initial opinions of different approaches on Cora. **c** Analysis of parameters in different datasets

Table 2 Comparison of fairness metrics in ablation study on fairness constraints (AV, RE, Cora, Twitter)

	AV				RF				Cora				Twitter			
Method	12	20	28	36	12	20	28	36	12	20	28	36	12	20	28	36
w/o Fairness	0.05	0.07	0.08	0.07	0.04	0.04	0.05	0.09	0.06	0.07	0.07	0.07	0.04	0.03	0.07	0.05
OMBF	0.08	0.11	0.12	0.12	0.11	0.16	0.21	0.25	0.14	0.18	0.21	0.23	0.07	0.16	0.21	0.18

These values (12, 20, 28, 36) represent the seed budget size

opinions to compare with benchmarks in Cora dataset. From Fig. 7a, b, we can see that both active nodes opinions and fair values in OMBF are higher than other approaches which is consistent with the results of randomly setting initial opinions in Cora dataset. In specific cases and under the condition of random assignment about initial opinions, the proposed OMBF shows better results than the other three approaches, indicating the effectiveness of the proposed OMBF to a certain extent. In summary, OMBF not only ensures the fairness of minority groups, but also ensures the good reputation of the target product that is the goal of OM problem.

5.3.6 Ablation Study

To rigorously evaluate the OMBF framework, we conduct ablation studies on two critical aspects: (1) the role of fairness constraints and (2) the impact of β. First, removing the fairness constraint (w/o Fairness) leads to a substantial decline in fairness metrics for minority groups. As shown in Table 2, on the Antelope-Valley dataset, the fairness value drops by 4.5%, confirming that fairness constraints are indispensable for equitable dissemination. Second, we explore the sensitivity of β across datasets. As shown in Fig. 7c, each dataset exhibits an optimal β value that maximizes the total opinions of active nodes. For instance, on Cora, $\beta = 0.013$ achieves the highest total opinions (40.7), while $\beta = 0.025$ reduces opinions by 12%. Similar patterns are observed in other datasets (e.g., $\beta = 0.009$ for Rice-Facebook). These results highlight β's dual role as a fairness regulator and a tunable parameter for opinion optimization.

6 Conclusions

In this research, we formulate the fairness of information dissemination in the opinion maximization problem in minority groups. Different from previous studies, we design OMBF to solve the OM problem, including three parts: (1) the determination of candidate nodes using node representations by GAT, (2) the dynamic change in opinions and (3) the selection of final seed nodes. For the first phase, we use GAT to obtain a high quantity of node representations in low dimensional space, which could extract the network structural information well, and further design an algorithm to determinate candidate nodes in the embedding space. For the second phase, we design the opinion propagation model (OLT) to propagate information and according to the opinion formation model to dynamically change node opinions, including active and inactive nodes. For the third phase, we select the final seed set for each group in the candidate set according to the seed budget and the fair value which is to ensure the fairness in the minority group. Finally, to demonstrate the effectiveness of our proposed algorithm OMBF, we compared it with three baseline methods and Greedy method on synthetic and real-world social networks. Experimental results show that our proposed OMBF has superior opinions in active nodes and fair values than the chosen benchmarks. Furthermore, in terms of overall opinion, our algorithm could obtain similar total positive opinions compared to Greedy method but with much lower computational complexity.

Acknowledgements This work was supported in part by the National Key Research and Development Program of China under Grant No. 2024YFF0617700; the National Natural Science Foundation of China (Nos. U22A2025), the Fundamental Research Funds for the Central Universities No. N2216011.

Author contributions Yingying Zhai involved in conceptualization, methodology, investigation, data curation, writing (original draft preparation and reviewing), and funding (No. 2024YFF0617700,N2216011). Yuting Lin involved in investigation, validation, data curation, software, visualization, and writing , Zhenling Han, original draft preparation and reviewing. Zefang Dong involved in conceptualization and writing (reviewing and editing). Xiaochun Yang involved in conceptualization, writing (reviewing and editing) and funding (No. U22A2025). Bin Wang, involved in conceptualizationfunding and (Nos. 2024YFF0617700, 62072088), Yuliang Cai involved in reviewing and editing.

Data availability The data are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that there are no financial and personal relationships with other people or organizations that can inappropriately influence our work. And there are no potential Conflict of interest with respect to this work.

References

1. Nguyen HT, Thai MT, Dinh TN (2017) A billion-scale approximation algorithm for maximizing benefit in viral marketing. IEEE/ACM Trans Netw (TON) 25(4):2419–2429. https://doi.org/10.1109/TNET.2017.2691544
2. Tang S (2018) When social advertising meets viral marketing: sequencing social advertisements for influence maximization. In: Thirty-second AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v32i1.11306
3. Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining, pp 420–429. https://doi.org/10.1145/1281192.1281239
4. Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: ACM SIGKDD international conference on knowledge discovery and data mining . https://doi.org/10.1145/1557019.1557047
5. Li Y, Fan J, Wang Y, Tan K-L (2018) Influence maximization on social graphs: a survey. IEEE Trans Knowl Data Eng 30(10):1852–1872. https://doi.org/10.1109/TKDE.2018.2807843
6. Lee JK, Choi J, Kim C, Kim Y (2014) Social media, network heterogeneity, and opinion polarization. J Commun 64(4):702–722. https://doi.org/10.1111/jcom.12077
7. Ellison NB, Steinfield C, Lampe C (2007) The benefits of facebook friends: social capital and college students' use of online social network sites. J Comput Mediat Commun 12(4):1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x
8. Hosseinalipour S, Nayak A, Dai H (2017) Real-time strategy selection for mobile advertising in VANETs. In: 2017 IEEE global communications conference, pp 1–6 (GLOBECOM 2017). https://doi.org/10.1109/GLOCOM.2017.8253975
9. Nayak A, Hosseinalipour S, Dai H (2017) Dynamic advertising in VANETs using repeated auctions. In: 2017 IEEE global communications conference (GLOBECOM 2017), pp 1–6. https://doi.org/10.1109/GLOCOM.2017.8254662
10. He Q, Zhang D, Wang X, Ma L, Zhao Y, Gao F, Huang M (2022) Graph convolutional network-based rumor blocking on social networks. IEEE Trans Comput Soc Syst. https://doi.org/10.1109/GLOCOM.2017.8253975
11. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, pp 137–146. https://doi.org/10.1145/956750.956769
12. Goya, A, Lu W, Lakshmanan LVS (2011) Celf++: optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th international conference companion on world wide web, pp 47–48. https://doi.org/10.1145/1963192.1963217
13. Samadi M, Nagi R, Semenov A, Nikolaev A (2018) Seed activation scheduling for influence maximization in social networks. Omega 77:96–114. https://doi.org/10.1016/j.omega.2017.06.002
14. Liu X, Kong X, Yu PS (2018) Active opinion maximization in social networks. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1840–1849. https://doi.org/10.1145/3219819.3220061
15. Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks.

In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1029–1038. https://doi.org/10.1145/1835804.1835934
16. Chen W, Collins A, Cummings R, Ke T, Liu Z, Rincon D, Sun X, Wang Y, Wei W, Yuan Y (2011) Influence maximization in social networks when negative opinions may emerge and propagate. In: Proceedings of the 2011 SIAM international conference on data mining (SDM), pp 379–390. https://doi.org/10.1137/1.9781611972818.33
17. Wang H, Yang Q, Lei F, Lei W (2015)Maximizing positive influence in signed social networks. In: SIGCHI conference on human factors in computing systems. https://doi.org/10.1007/978-3-319-27051-7_30
18. Gionis A, Terzi E, Tsaparas P (2013) Opinion maximization in social networks. In: Proceedings of the 2013 SIAM international conference on data mining (SDM), pp 387–395. https://doi.org/10.1137/1.9781611972832.43
19. Abebe R, Kleinberg J, Parkes D, Tsourakakis CE (2018) Opinion dynamics with varying susceptibility to persuasion. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery and data mining, pp 1089–1098. https://doi.org/10.1145/3219819.3219983
20. Wen Z, Kveton B, Valko M, Vaswani S (2017) Online influence maximization under independent cascade model with semi-bandit feedback. Adv Neural Inf Process Syst 30:3022–3032
21. Ahmadinejad A, Dehghani S, Hajiaghayi M, Mahini H, Seddighin S, Yazdanbod S (2014) How effectively can we form opinions? In: Proceedings of the 23rd international conference on world wide web, pp 213–214. https://doi.org/10.1145/2567948.2577201
22. Chen Y, Li H, Qu Q (2018) Negative-aware influence maximization on social networks. Proceedings of the AAAI Conference on Artificial Intelligence. https://doi.org/10.1609/aaai.v32i1.12149
23. Han K, Xu C, Gui F, Tang S, Huang H, Luo J (2018) Discount allocation for revenue maximization in online social networks. In: Proceedings of the eighteenth ACM international symposium on mobile ad hoc networking and computing, pp 121–130. https://doi.org/10.1145/3209582.3209595
24. Liu Q, Xiang B, Chen E, Xiong H, Tang F, Yu JX (2014) Influence maximization over large-scale social networks: A bounded linear approach. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management, pp 171–180. https://doi.org/10.1145/2661829.2662009
25. Zhang H, Dinh TN, Thai MT (2013) Maximizing the spread of positive influence in online social networks. In: 2013 IEEE 33rd international conference on distributed computing systems, pp 317–326 . https://doi.org/10.1109/ICDCS.2013.37
26. Borgs C, Brautbar M, Chayes J, Lucier B (2014) Maximizing social influence in nearly optimal time. In: Proceedings of the twenty-fifth annual ACM–SIAM symposium on discrete algorithms, pp 946–957. https://doi.org/10.1137/1.9781611973402.70
27. Keikha MM, Rahgozar M, Asadpour M, Abdollahi MF (2020) Influence maximization across heterogeneous interconnected networks based on deep learning. Expert Syst Appl 140:112905. https://doi.org/10.1016/j.eswa.2019.112905
28. Panagopoulos G, Malliaros FD, Vazirgiannis M (2022) Multitask learning for influence estimation and maximization. IEEE Trans Knowl Data Eng 34(9):4398–4409. https://doi.org/10.1109/TKDE.2020.3040028
29. Tian S, Mo S, Wang L, Peng Z (2020) Deep reinforcement learning-based approach to tackle topic-aware influence maximization. Data Sci Eng 5(1):1–11. https://doi.org/10.1007/s41019-020-00117-1
30. Keikha MM, Rahgozar M, Asadpour M (2018) Community aware random walk for network embedding. Knowl-Based Syst 148:47–54. https://doi.org/10.1016/j.knosys.2018.02.028
31. Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. https://doi.org/10.48550/arXiv.1301.3781
32. Zhang C, Li W, Wei D, Liu Y, Li Z (2022) Network dynamic GCN influence maximization algorithm with leader fake labeling mechanism. IEEE Trans Comput Soc Syst. https://doi.org/10.1109/TCSS.2022.3193583
33. He Q, Wang X, Lei Z, Huang M, Cai Y, Ma L (2019) TIFIM: a two-stage iterative framework for influence maximization in social networks. Appl Math Comput 354:338–352. https://doi.org/10.1016/j.amc.2019.02.056
34. Fish B, Bashardoust A, Boyd D, Friedler S, Scheidegger C, Venkatasubramanian S (2019) Gaps in information access in social networks? In: The world wide web conference, pp 480–490. https://doi.org/10.1145/3308558.3313680
35. Rahmattalabi A, Vayanos P, Fulginiti A, Rice E, Wilder B, Yadav A, Tambe M (2019) Exploring algorithmic fairness in robust graph covering problems. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R (eds) Advances in neural information processing systems, vol 32, Curran Associates, Inc., Red Hook, NY, USA, pp 15750–15761.
36. Ali J, Babaei M, Chakraborty A, Mirzasoleiman B, Gummadi KP, Singla A (2023) On the fairness of time-critical influence maximization in social networks. IEEE Trans Knowl Data Eng 35(3):2875–2886. https://doi.org/10.1109/TKDE.2021.3120561
37. Khajehnejad M, Rezaei AA, Babaei M, Hoffmann J, Jalili M, Weller A (2020) Adversarial graph embeddings for fair influence maximization over social networks. In: Proceedings of the twenty-ninth international joint conference on artificial intelligence (IJCAI ’20), pp 4306–4312. https://doi.org/10.48550/arXiv.2005.04074
38. Khajehnejad A, Khajehnejad M, Babaei M, Gummadi KP, Weller A, Mirzasoleiman B (2022) Crosswalk: fairness-enhanced node representation learning, vol 36, pp 11963–11970. https://doi.org/10.1609/aaai.v36i11.21454
39. He Q, Fang H, Zhang J, Wang X (2023) Dynamic opinion maximization in social networks. IEEE Trans Knowl Data Eng 35(1):350–361. https://doi.org/10.1109/TKDE.2021.3077491
40. He Q, Wang X, Zhao Y, Yi B, Lu X, Yang M, Huang M (2021) Reinforcement-learning-based competitive opinion maximization approach in signed social networks. IEEE Trans Comput Soc Syst. https://doi.org/10.1109/TCSS.2021.3120421
41. He Q, Lv Y, Wang X, Li J, Huang M, Ma L, Cai Y (2022) Reinforcement learning based dynamic opinion maximization framework in signed social networks. IEEE Trans Cogn Dev Syst. https://doi.org/10.1109/TCDS.2022.3141952
42. Khajehnejad M (2019) SimNet: similarity-based network embeddings with mean commute time. PLoS ONE 14(8):0221172. https://doi.org/10.1371/journal.pone.0230096
43. Grover A, Leskovec J (2016) Node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 855–864 . https://doi.org/10.1145/2939672.2939754
44. Wang D, Cui P, Zhu W (2016) Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1225–1234. ACM. https://doi.org/10.1145/2939672.2939753
45. Li H, Xu M, Bhowmick SS, Sun C, Jiang Z, Cui J (2019) Disco: influence maximization meets network embedding and deep learning. arXiv preprint arXiv:1906.07378. https://doi.org/10.48550/arXiv.1906.07378
46. Hamilton WL, Ying R, Leskovec J (2017) Inductive representation learning on large graphs. In: Proceedings of the 31st international conference on neural information processing systems, pp 1025–1035. Curran Associates Inc., Red Hook, NY, USA

47. Velikovi P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2017) Graph attention networks. https://doi.org/10.48550/arXiv.1710.10903
48. Mislove A, Viswanath B, Gummadi PK, Druschel P (2010) You are who you know: inferring user profiles in online social networks. In: Proceedings of the third international conference on web search and web data mining (WSDM 2010), February 4–6, 2010, New York, NY, USA. https://doi.org/10.1145/1718487.1718519
49. Wilder B, Ou HC, Haye K, Tambe M (2018) Optimizing network structure for preventative health. In: Proceedings of the 17th international conference on autonomous agents and multiagent systems, pp 841–849
50. Cora Network Dataset-RELATIONAL (2008). https://relational.fit.cvut.cz/dataset/CORA
51. Tsang A, Wilder B, Rice E, Tambe M, Zick Y (2019) Group-fairness in influence maximization. arXiv preprint arXiv:1903.00967
52. Mokhtarizadeh S, Zamani Dehkordi B, Mosleh M, Barati A (2021) Influence maximization using time delay based harmonic centrality in social networks. Tabriz J Electr Eng 51(3):359–370
53. Tsang A, Wilder B, Rice E, Tambe M, Zick Y (2019) Group-fairness in influence maximization. arXiv preprint arXiv:1903.00967
54. Regulation P (2018) General data protection regulation. InTouch 25:1–5
55. Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. Soc Netw 32(3):245–251. https://doi.org/10.1016/j.socnet.2010.03.006
56. Natenzon P (2018) Random choice and learning. J Polit Econ. https://doi.org/10.1086/700762
57. Alla LS, Kare AS (2023) Opinion maximization in signed social networks using centrality measures and clustering techniques. In: International conference on distributed computing and intelligent technology, pp 125–140. Springer
58. Wang J, Cao Z, Xie C, Li Y, Liu J, Gao Z (2025) DGN: influence maximization based on deep reinforcement learning. J Supercomput 81(1):130