Networks and Heterogeneous Media

NHM, 21 (7): 2468–2503.
DOI: 10.3934/nhm.2026092
Received: 15 September 2026
Revised: 14 November 2026
Accepted: 17 November 2026
Published: 25 November 2026

https://www.aimspress.com/journal/NHM

Research article

Integrating game theory and reinforcement learning for optimizing truth source selection strategies: A time-sensitive influence maximization method in signed social networks

Ying Guo and Yong-nan Li*

School of National Security, People's Public Security University of China, Beijing 100038, China

* **Correspondence:** Email: liyongnan@ppsuc.edu.cn.

Abstract: With the advent of the data-intelligence era, the risk of rumor propagation on online social networks is increasingly threatening the security of online ideology. By deploying the truth to capture the high ground of public opinion and create a favorable ecological environment for online public opinion, the breeding ground of rumors can be thoroughly eradicated, and the discourse power of online public opinion can be firmly grasped. Therefore, this paper focuses on the problem of time-sensitive dominant competitive influence maximization. Without knowing the opponent's strategy, a truth source selection algorithm for directed signed social networks is proposed to gain greater competitive advantages, seize the strategic dominant position, and proactively shape an overall favorable trend for future public opinion development. First, the structural balance theory and the echo chamber effect are utilized to optimize the dissemination rules, and a continuous opinion competition propagation model suitable for directed signed social networks is proposed. Second, by using directed signed graph embedding and mixed-strategy Nash equilibrium, the state space and action space are reconstructed, and a truth source selection algorithm integrating game theory and double deep Q-network is proposed. Finally, experiments are conducted on four real social network datasets. The results show that the algorithm proposed in this paper further improves the quality of the positive seed set compared with state-of-the-art algorithms and has good robustness and interpretability. Meanwhile, in addition to the placement of truth sources, this paper also discusses the effects of the number and duration of their deployment on intervention outcomes, thereby forming a comprehensive deployment plan for truth sources and providing actionable practical guidance for strategic decision-making in rumor governance.

Keywords: time-sensitive dominant competitive influence maximization; mixed-strategy Nash equilibrium; double deep Q-network; directed signed social network; continuous opinion competition propagation model

1. Introduction

The spread of online rumors not only causes harm to individuals and society but also poses a serious threat to national security. Strategies for governing online rumors are generally categorized into two types. First, locating rumor sources to target the driving force of dissemination and block the rumor evolution process [1]. Second, screening truth sources to suppress rumor spread with facts and capture the dominant position in online public opinion. From the perspective of systemic governance, the governance of online rumors must be holistic: it should address symptoms in the short term while also addressing root causes in the long term. By strategically positioning sources of truth online, truth can gain the upper hand in its competition with rumors, proactively shape the future trajectory of public opinion, and prevent the emergence and spread of rumors at the source. Consequently, the problem of influence maximization has increasingly become one research hotspot. The traditional influence maximization (IM) problem refers to finding a set of seed nodes that maximizes the scope of influence propagation. Currently, the variants of the IM problem are primarily classified into the following categories:

First, in terms of influence polarity, the IM problem can be categorized into positive influence maximization [2], negative influence minimization [3], and net positive influence maximization [4]. Zhu et al. [2] addressed the positive influence maximization (PIM) problem in signed social networks by proposing the polar and decay related independent cascade (IC-PD) model and designing a signed reverse influence sampling (SRIS) algorithm. This algorithm transforms the influence maximization problem into the set cover problem, selecting the node set that covers the largest number of polar reverse reachable (PRR) sets as the optimal seed set. However, this sampling-based approach requires storing a large number of sets, imposing high demands on storage resources. Dong et al. [3] focused on the rumor blocking maximization (RBM) problem in unsigned networks, introducing a competitive linear threshold model with state transition (CLTST) and a community-based rumor blocking (CRB) algorithm. The algorithm selects and activates the anti-rumor seed set to minimize the number of nodes ultimately activated by rumors; first, by evaluating the community influence of the node to select candidate anti-rumor seeds; and second, employing an adaptive search strategy to obtain the optimal anti-rumor seed set. Li et al. [4] proposed the net positive influence maximization (NPIM) problem in signed social networks, which aims to select a positive seed set to maximize final positive influence while minimizing negative influence. First, a polarity-related independent cascade (IC-P) model was developed for single information dissemination. Second, an improved R-greedy algorithm was designed, which accounts for the negative influence exerted by positive seeds via negative edges and optimizes the greedy framework to reduce runtime. However, the efficiency of this traditional algorithm remains relatively low.

Second, in terms of the number of sources, the IM problem can be categorized into single information dissemination [5–8] and competitive information dissemination. Competitive information dissemination refers to the scenario where positive and negative information propagate in competition, which is further subdivided into dominant competition [9] and dominant protection [10]. The former sets an upper limit on the number of positive seeds to maximize positive influence; the latter sets a

lower limit on positive influence to minimize the number of positive seeds. Zhu et al. [8] addressed the influence maximization (IM) problem in single information dissemination by proposing an end-to-end framework, BiGDN, based on deep reinforcement learning (DRL) and graph neural networks (GNNs). This framework consists of two core components: the node encoder (NE) and the Q-network (QN). First, in the NE component, GNN and a bidirectional neighborhood aggregation strategy are employed to learn the initial embeddings of nodes. Second, in the QN component, combined state convolutional layer and a multi-head attention mechanism are used to learn the final node embeddings and the parameters of the Q-network. Finally, experimental results validate that the learning-based method achieves a balance between efficiency and effectiveness and exhibits superior generalization ability across most datasets. Li et al. [9] proposed the dominated competitive influence maximization (DCIM) problem in unsigned social networks, which aims to select a positive seed set to maximize the influence of target information while minimizing the influence of competing information. First, considering time limitation and time delay, a competitive independent cascade model with meeting events (CIC-M) was developed. Second, the DCIM_CELF algorithm was designed, which is based on the cost-effective lazy forward (CELF) algorithm and leverages the submodularity of the propagation function to prune candidate nodes. However, the time complexity of the optimized greedy algorithm remains relatively high.

Third, from the perspective of cost-benefit, the IM problem can be categorized into minimum cost [11] and maximum benefit. The former aims to achieve the expected influence range with the minimum number of seeds, while the latter pursues influence maximization under a constrained number of seeds. Maximum benefit is further subdivided into maximization of total opinion value [12–14] and maximization of the number of affected nodes [15–17]. The former quantifies the influence of the seed set as the sum of opinions in the network, whereas the latter quantifies it as the total count of affected nodes. Yu et al. [13] investigated the opinion maximization (OM) problem, which targets selecting a positive seed set to maximize the total opinion value in the network. First, they proposed an opinion dynamic model based on game theory, treating individual stubbornness as a strategy choice. The strategy was first updated through social interaction, followed by synchronous opinion update based on the current strategy, thereby modeling the opinion propagation process. Second, they designed a two-stage opinion maximization algorithm based on deep Q-network (DQN), namely T-DQN. The first stage leverages heuristic rules to select potential seeds, while the second stage employs DQN to identify the optimal seed set from potential candidates. He et al. [14] addressed the competitive opinion maximization (COM) problem in social networks, considering the scenario of unknown opponent's strategy. First, they proposed an activated dynamic opinion model based on stateless Q-learning to simulate the opinion interaction process between nodes. Second, they developed a brand-new Q-learning-based opinion maximization framework (QOMF), which effectively reduces the complexity of the action space by designing four simple seed selection strategies. However, compared to deep Q-learning, Q-learning struggles to adapt to high-dimensional continuous state spaces and fails to handle more complex learning tasks.

Fourth, from the perspective of time sensitivity, the IM problem can be categorized into limited time (i.e., time-sensitive) [18,19] and termination time (i.e., the time when node states are stable) [20]. The former integrates temporal factors into the decision-making framework to maximize influence within a limited time, while the latter aims to maximize influence upon the termination of propagation. Wang et al. [18] integrated the factors affecting influence diffusion into influence maximization, considering the timeliness of influence diffusion and polarity of influence itself, and proposed the

time-sensitive positive influence maximization (TP-IM) problem. The goal of TP-IM is to select the positive seed set that maximizes positive influence within a specified time limit. To address this, they first proposed the heat diffusion–based polarity influence diffusion (HDPID) model for signed social networks. Second, they designed an improved k-step greedy seed node selection (TP-IM k-step greedy) algorithm tailored to solve the TP-IM problem. He et al. [19] formulated the positive influence maximization in signed networks within a limited time (PIMST) problem, aiming to find the positive seed set so that the positive influence spreads through paths containing only positive edges and reaches the maximum range with time limit. Their approach involves two key steps: The first is combining two heuristic methods to generate candidate seed nodes; the second is proposing the forward and inverted index cross lazy forward (FICLF) algorithm based on CELF, significantly reducing the time complexity of the classical greedy algorithm. Sung et al. [20] defined a signed-aware influence maximization (SIM) problem, which seeks an optimal seed set to simultaneously maximize positive influence and minimize negative influence at the end of the propagation process. First, by incorporating binary opinions and signed relationships into the independent cascade (IC) model, they proposed a symbol-aware competitive independent cascade (SCIC) model that accounts for positive as well as negative dominance mechanisms in the influence propagation process. Second, they proposed the signed-aware competitive maximum influence arborescence (S-CMIA) algorithm to select the optimal positive seed set, though this method does not yield an approximate solution with theoretical guarantees.

Fifth, in terms of method categories, the IM problem can be divided into traditional methods and machine learning–based methods [21]. Traditional methods mainly include approximation-based methods and heuristic-based methods, while machine learning–based methods primarily encompass graph representation learning-based and deep reinforcement learning–based. For approximation-based methods, if the propagation function satisfies submodularity and monotonicity, the greedy algorithm can achieve a theoretically guaranteed approximation ratio of at least $(1 - 1/e)$. Such methods exhibit high accuracy but low computational efficiency. To further reduce time costs, heuristic algorithms emerged, which feature lower time complexity but fail to guarantee approximation accuracy in the worst case. In contrast, machine learning–based methods strike a favorable balance between accuracy and computational efficiency. Different from traditional approaches, the DRL framework offers a novel solution paradigm. It regards the optimal seed selection as a combinatorial optimization problem, treats the network state as an overall environment, and defines the action as selecting an optimal seed. The agent learns a reward function for state–action pairs through continuous interaction with the environment and feedback from experience. It engages in incremental learning and iterative optimization, selecting actions that maximize the expected long-term reward based on the current state. Ultimately, a decision sequence is generated to constitute the optimal seed set.

Recent studies have integrated graph representation learning with deep reinforcement learning to tackle the IM problem in social networks. Li et al. [22] presented the PIANO framework that incorporates network embedding and DQN, which outperforms traditional methods in terms of runtime without sacrificing seed quality and avoids the costly propagation sampling. Chen et al. [23] proposed the ToupleGDD framework composed of three coupled GNNs and double DQN (DDQN), demonstrating strong generalization ability on synthetic and realistic datasets. Wang et al. [24] proposed the DGN framework, which combines dual coupled GNNs and DQN, showing excellent robustness in large-scale complex social networks.

It should be noted that IM problems can also be classified from the perspective of diffusion models.

In terms of describing the information diffusion process, the main categories are influence models and opinion models. The former involves random diffusion through influence activation, with node states typically classified as activated or inactive; the latter involves consensus evolution through opinion update mechanisms, with node states typically represented as discrete or continuous opinion values, providing a finer-grained characterization of node states. Kempe et al. [25] formulated the IM problem as a discrete optimization problem and proposed a general threshold and cascade diffusion model. Through computational experiments on large-scale collaboration networks, they demonstrated that the greedy algorithm significantly outperforms heuristic approaches based on degree centrality and distance centrality. Yildiz et al. [26] constructed a binary opinion dynamics model with stubborn individuals (i.e., individuals who hold fixed opinions and have strong influence). In this model, stubborn individuals holding opposing opinions hinder the achievement of social consensus. They also proposed a greedy algorithm to solve the optimal placement problem for stubborn individuals by selecting their optimal positions to maximize their influence on other individuals. Vassio et al. [27] designed a distributed message-passing algorithm based on an opinion dynamics model with stubborn individuals to compute the harmonic influence centrality (HIC) they proposed. Heuristic methods based on HIC values can solve the optimal leader selection problem, maximizing the change in the average opinion of a social network. Bini et al. [28] addressed the optimal target selection problem in competitive dynamics models by constructing more refined heuristics based on the underlying graph structure. In the competition between two strategic individuals holding opposing opinions, these heuristics maximize the spread of their own opinions while blocking the opponent's influence, allowing their own opinions to dominate, and thereby steering the average asymptotic opinion. Zhang et al. [29] proposed an uncertainty-based opinion model and used deep reinforcement learning to solve the competitive influence maximization problem. The model uses subjective logic to model uncertainty in users and decision-making, enabling true information to retain influence even in the face of sophisticated disinformation strategies. Chang et al. [30] modeled influence propagation under binary competition as a flow process, estimating the power of influence flow via edge-view. They also proposed a scalable heuristic that exploits polarity reversal in signed networks to simultaneously enhance positive influence and weaken negative influence. All of the above problems are essentially variants of the IM problem, and stubborn individuals, strategic individuals, opinion leaders, and key nodes [31] are equivalent to the positive seed set.

As a summary, a comparison of relevant studies on the IM problem is provided in Table 1. As can be seen intuitively from Table 1, the research objective of this paper differs from that of existing IM variants in terms of the problem, at least in one classification dimension; therefore, new algorithms should be customized according to the specific characteristics of the problem.

Table 1. Comparison table of relevant algorithms.

Year	Algorithm	Problem	Network	Source	Time	Method
2023	SRIS [2]	PIM	Signed	Single	Sensitive	Traditional
2024	CRB [3]	RBM	Unsigned	Competitive	Insensitive	Traditional
2021	Improved R-greedy [4]	NPIM	Signed	Single	Insensitive	Traditional
2025	BiGDN [8]	IM	Unsigned	Single	Sensitive	Learning
2018	DCIM_CELF [9]	DCIM	Unsigned	Competitive	Sensitive	Traditional
2026	T-DQN [13]	OM	Signed	Single	Sensitive	Learning
2025	QOMF [14]	COM	Unsigned	Competitive	Sensitive	Learning
2021	TP-IM k-step greedy [18]	TP-IM	Signed	Single	Sensitive	Traditional
2023	FICLF [19]	PIMST	Signed	Single	Sensitive	Traditional
2020	S-CMIA [20]	SIM	Signed	Competitive	Insensitive	Traditional
2023	PIANO [22]	IM	Unsigned	Single	Sensitive	Learning
2024	ToupleGDD [23]	IM	Unsigned	Single	Sensitive	Learning
2025	DGN [24]	IM	Unsigned	Single	Sensitive	Learning
Ours	DM-DDQN	TDCIM	Signed	Competitive	Sensitive	Learning

For the IM problem in online social networks, existing studies predominantly focused on scenarios of single information diffusion, single paths consisting of positive edges, and the maximization of net positive influence. In contrast, research that simultaneously accounts for the competitive diffusion of positive and negative information, mixed paths composed of both positive and negative edges, time as a sensitive factor, and the maximization of dominant competitive influence remains relatively scarce. Notably, this problem is conducive to formulating strategies for rumor governance, precisely deploying truth sources to curb rumor propagation, and creating a clean and upright online public opinion environment. Therefore, this paper focuses on optimal strategy selection under the scenario of unknown opponent's strategy. First, a fine-grained opinion propagation model is constructed in the directed signed social network. Second, by balancing exploration and exploitation, as well as immediate and delayed rewards, the optimal positive seed set within a finite time is obtained to achieve strategy optimization in a competitive environment. Our key contributions are outlined as follows:

(1) This paper, in response to the practical demands of online rumor governance, proposes the time-sensitive dominant competitive influence maximization (TDCIM) problem. Subsequently, drawing on the characteristics of this problem, the dissemination model is optimized, and the seeding algorithm is customized. Finally, experiments are carried out on real-world online social networks to analyze the impact of the quantity and duration of truth sources on dissemination outcomes, thereby providing candidate truth source schemes and expected governance effects for the formulation of rumor control measures.

(2) This paper optimizes the rules for opinion updating and competition and proposes a continuous opinion competition (COC) propagation model that integrates the echo chamber effect and structural balance theory, which is applicable to rumor propagation scenarios in real-world online social networks. First, the weights of out-degree neighbors' opinions are jointly determined using the degree of consensus and the degree of trust. Second, the signs of out-degree neighbors' opinions are determined based on the type of trust relationship. Finally, the final COC model is derived based on the Friedkin–Johnsen model.

(3) This paper introduces game theory to optimize the decision-making framework and proposes an improved DDQN algorithm that integrates directed signed graph embedding and mixed-strategy Nash equilibrium (DM-DDQN), enabling the selected positive seed set to gain more competitive advantages and strategic initiative. First, leveraging the signed graph attention mechanism, we improve the signed directed graph neural networks (SDGNN) for node embeddings and reconstruct the state space using the sum of embeddings of positive and negative seeds. Second, in the context of unknown opponent's strategy, we solve for the mixed-strategy Nash equilibrium and reconstruct the action space based on the probability distribution of four actions. Finally, based on the ε-greedy strategy and the DDQN algorithm, we derive the final DM-DDQN algorithm, which is employed to select the optimal positive seed set.

The current research faces a dilemma: Under the classic IC and LT models, the IM problem is NP-hard [25], and variants of the IM problem under extended IC and LT models are generally also NP-hard. At the same time, for the CIM problem, current public datasets only provide the true propagation graphs of rumors (e.g., PHEME-9, Twitter16), lacking true competitive propagation graphs of truth and rumors within the same network and overlooking the opponent's countermeasures and our own defenses under strategic interaction between the two sides. Therefore, most studies constructed information propagation models that closely reflect real-world scenarios based on real social network datasets and conducted competitive propagation experiments through simulation to trace the evolution of opinions and validate the superiority of approximate solutions. However, in terms of propagation modeling, few studies integrated sociological theories to optimize propagation rules; in terms of seed selection strategies, few studies combined the complementary strengths of game theory and reinforcement learning.

Therefore, the unique contribution of this paper lies in proposing a new variant of the IM problem and customizing a model and algorithm tailored to it, based on the practical needs of online rumor governance. First, this paper constructs the COC model using social psychology theories, making it more aligned with real-world scenarios in which rumors and truth compete for propagation. Second, this paper jointly optimizes Nash equilibrium and DDQN and employs prioritized experience replay and soft updates to improve training stability. The Nash equilibrium strategy is used to select the current optimal strategy, while DDQN estimates the discounted long-term cumulative reward, allowing the two to complement each other. Meanwhile, in response to the characteristics of the TDCIM problem, this paper reconstructs the state space and action space of DDQN, reflecting a high degree of synergy between the strategy and the task. Finally, experiments on real-world social network datasets demonstrate that the truth source placement strategy proposed in this paper has significant advantages, illustrating the adaptability of the model to the scenario and the algorithm to the problem.

The overall process of the DM-DDQN algorithm proposed in this paper is illustrated in Figure 1, which consists of two interactive components: seed selection and parameter update. In the seed selection phase, under the scenario where the opponent's strategy is unknown, the ε-greedy strategy and mixed-strategy Nash equilibrium are employed to balance local optimality and global optimality, yielding an optimal positive seed and a new experience tuple. In the parameter update phase, three sequential steps are executed as follows: A batch of experiences is sampled with priority to improve sample utilization efficiency and accelerate convergence; the temporal difference (TD) error of the experiences is leveraged to construct the mean squared error (MSE) loss, which balances one-step optimality and two-step optimality; the parameters of the online network, θ_{online}, are updated via stochastic gradient descent (SGD), while a soft update is performed on the parameters of the target

network, θ_{target}. These two phases are executed alternately, with k positive seeds selected sequentially, so as to maximize the expected cumulative reward of the positive seed set.

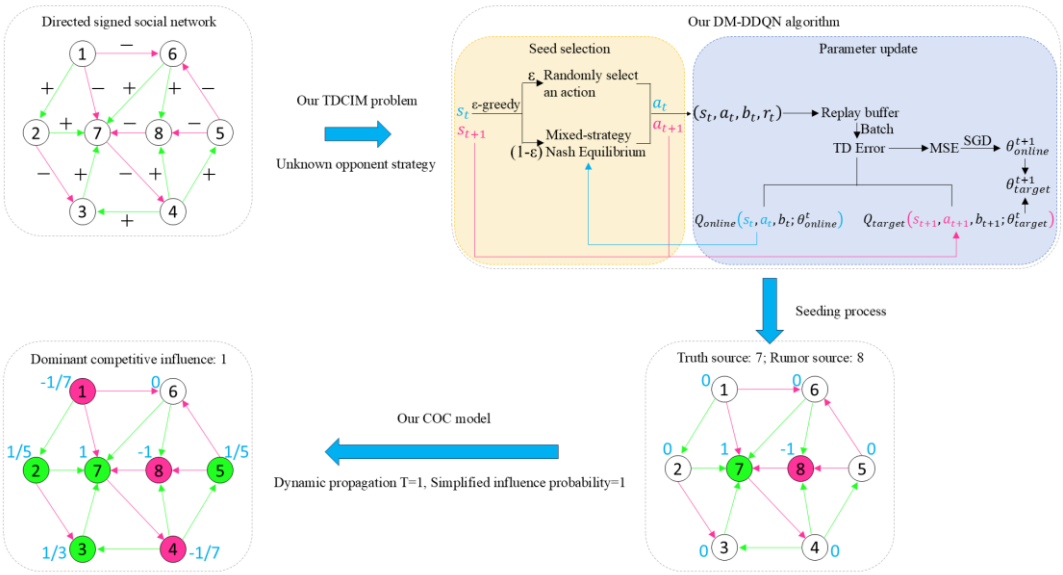

Figure 1. The overall process of truth source selection.

The structure of this paper is organized as follows: Section 2 presents the mathematical formulation of the TDCIM problem and concurrently constructs the COC propagation model. Section 3 elaborates in detail on the DM-DDQN seeding algorithm, encompassing the state space, action space, reward function, and training procedure. Section 4 assesses the superiority and robustness of the DM-DDQN algorithm via extensive experiments, along with the contributions of its individual components. It also identifies the limitations of our study and highlights avenues for future work. Section 5 concludes the paper.

2. Problem modeling

2.1. Problem definition

The time-sensitive dominant competitive influence maximization (TDCIM) problem is defined as follows: In a directed signed network $G(V,E)$ under a competitive setting where positive and negative seeds are deployed simultaneously, the goal is to select k positive seed nodes from the node set V such that the difference between the number of positive nodes and negative nodes is maximized after a specified time step T, as formalized in Eq (2.1). Here, positive nodes denote those holding positive opinions, and negative nodes denote those holding negative opinions.

$$S_*^+ = \underset{S^+\subseteq V\setminus S^-,|S^+|=|S^-|=k}{argmax} \left(I_T\left(S^+\right)-I_T\left(S^-\right)\right), \tag{2.1}$$

where S_*^+ denotes the optimal positive seed set; S^+ denotes the positive seed set, (i.e., the set consisting of truth sources); S^- denotes the negative seed set, (i.e., the set consisting of rumor

sources); $I_T(\cdot)$ is the propagation function, which represents the influence scope of a seed set after a propagation duration of T; $I_T(S^+)$ and $I_T(S^-)$ denote the number of positive and negative opinions, respectively, in the network after the propagation time T. It should be noted that, since truth and rumor are relative concepts, this paper assumes that positive seeds represent sources of truth, and negative seeds represent sources of rumor.

Table 2. Mathematical notations.

Notation	Description
$G(V,E)$	A directed signed network, where V denotes the node set, and E denotes the edge set.
v_i	The node i.
e_{ij}	The sign of the directed edge from v_i to v_j takes values in the set {-1,0,1}, where -1 denotes a distrust relationship, 1 denotes a trust relationship, and 0 denotes the absence of a directed edge.
ω_{ij}	The weight of the directed edge from v_i to v_j, which denotes the degree of trust.
o_i^t	The opinion held by v_i at time t takes values in the interval [-1,1], where negative values denote negative opinions, zero denotes a neutral opinion, and positive values denote positive opinions.
z_i	The embedding of v_i.
c_{ij}^t	The opinion similarity between v_i and v_j at time t (i.e., the degree of consensus) takes values in the interval [0,1].
σ	The influence probability takes values in the interval (0,1); $(1-\sigma)$ denotes the stubbornness probability.
S_*^+	The optimal positive seed set.
k	The number of optimal seeds (i.e., the cardinality of the optimal seed set).
T	The constrained propagation time (i.e., the specified time step).
N_i^{in}	The set of v_i's in-degree neighbors.
N_i^{out}	The set of v_i's out-degree neighbors.
N_i^{in+}	The set of v_i's in-degree neighbors connected by a positive edge.
N_i^{in-}	The set of v_i's in-degree neighbors connected by a negative edge.
N_i^{out+}	The set of v_i's out-degree neighbors connected by a positive edge.
N_i^{out-}	The set of v_i's out-degree neighbors connected by a negative edge.
A_i^{in}	The set of v_i's positive in-degree neighbors.
B_i^{in}	The set of v_i's negative in-degree neighbors.
A_i^{out}	The set of v_i's positive out-degree neighbors.
B_i^{out}	The set of v_i's negative out-degree neighbors.
s_t	The state at time t.
a_t	The agent's action at time t.
b_t	The opponent's action at time t.
r_t	The reward received at time t.
μ_1	The size of the replay buffer.
μ_2	The total number of experiences in the replay buffer.
μ_3	The batch size.
p_i	The priority of experience i.
P_i	The sampling probability of experience i.
γ	The discount factor, which balances current and future rewards, takes values in the interval (0, 1].
τ	The soft update coefficient.
φ	The initial learning rate.

It is noteworthy that at the initial propagation time, the opinion value of positive seed nodes is +1, that of negative seed nodes is −1, and that of other nodes is 0. During the propagation process, the opinion values of seed nodes remain fixed. This is because in real-world online social network scenarios, ordinary users are highly susceptible to the influence of opinion leaders, whereas opinion leaders are barely influenced by ordinary users [32]. The mathematical notations employed in this paper are summarized in Table 2.

2.2. Propagation model

Existing competitive diffusion models are mostly the competitive independent cascade (CIC) model [33] and the competitive linear threshold (CLT) model [34]. Leveraging the complementary advantages of the CIC model and the Friedkin–Johnsen (F-J) model [35], this paper models node state activation and opinion updating in influence diffusion, integrating the dynamic evolution of opinions and the game process of competing opinions. A continuous opinion competition propagation (COC) model tailored to the TDCIM problem is proposed, which realizes the dynamic update of continuous opinions in the competitive diffusion of positive and negative opinions, thereby authentically capturing the evolution of individual opinions in online social networks. This model accounts for trust relationships and individual stubbornness and jointly determines the importance of neighboring opinions by leveraging consensus advantage and tie-breaking rule. Additionally, it infers the sign of out-degree neighbors' opinions based on the type of trust relationship. The specific optimization rules are as follows: First, the echo chamber effect is incorporated, where higher attention weights are allocated to out-degree neighbors with greater opinion similarity. Second, the structural balance theory is introduced, whereby congruent opinions are propagated to neighbors with a trust relationship, while opposing opinions are disseminated to neighbors with a distrust relationship. Finally, the proportional probability tie-breaking rule is adopted, where weights are assigned according to the proportion of positive/negative out-degree neighbors to quantify the trust level of different neighbors.

2.2.1. Dynamic update rules for continuous opinions

Social interaction embodies a unity of randomness and determinacy. Users exhibit a certain probability of stubbornness while pursuing the maximization of psychic income, which mirrors the opinion propagation process in real-world social networks [36]. Therefore, this paper retains the inherent bias term in the F-J model but replaces individual stubbornness with a stubbornness probability to reflect whether a node is influenced by its neighbors (i.e., whether it is activated). It also introduces the tie-breaking rule of the CIC model in competitive scenarios and replaces the fixed probabilities of positive and negative influence with proportional probabilities, making it suitable for opinion evolution in competitive environments. In addition, this paper introduces social psychology theories to optimize the propagation rules. Based on the echo chamber effect and structural balance theory, it leverages trust relationships and consensus advantages to quantify the magnitude and sign of neighbors' influence, thereby reflecting the information propagation characteristics of signed social networks and better aligning with rumor propagation scenarios in real-world social networks.

In the signed social network, first, based on the structural balance theory in social psychology [37,38], users tend to maintain similar views as those they trust and opposite views as those they distrust, while the strength of a relationship determines the importance of adjacent views. The method for calculating

the degree of trust is introduced in Section 2.2.2. Second, based on the echo chamber effect [39], in positive relationships, users are more likely to hear the echo of their own voices and thus satisfy the assimilating rule in opinion dynamics [40], tending to reach consensus with similar views. The more similar the views are, the stronger the positive attraction. Therefore, this paper models the consensus advantage of trusted neighbors as view similarity. Finally, based on the boomerang effect [41], in negative relationships, users are more likely to shift to opposing positions and thus satisfy the opposing rule [42], tending to reach consensus with opposite views. The greater the difference between views, the stronger the negative attraction. Therefore, this paper models the consensus advantage of untrusted neighbors as view dissimilarity. In summary, positive relationships attract identical views: The more similar the views are, the stronger the positive attraction, and the strength of assimilation is inversely proportional to the view distance. Conversely, negative relationships attract opposite views: The greater the difference between views, the stronger the negative attraction, and the strength of alienation is directly proportional to the view distance.

Therefore, this paper integrates the echo chamber effect and structural balance theory, leveraging the complementary advantages of the CIC model and F-J model. It also incorporates both the occurrence and magnitude of influence to enable continuous updates of opinions, modeling the evolution of individual opinions as a fusion of stubbornness and strategic behavior. Each node allocates probabilities to adhering to its original opinion or being influenced by others, while assigning weights to neighbors based on their importance. By optimizing opinion activation and update rules, an opinion propagation model tailored to online social networks is constructed. The implementation steps are as follows: First, the stubbornness probability is employed to determine whether a node is influenced by its neighbors. Second, if influenced, attention weights are assigned to neighbors' opinions to update its own opinion. Guided by the echo chamber effect and tie-breaking rule, the degree of neighbors' influence is jointly determined by the level of consensus and trust. Finally, based on structural balance theory, the sign of neighbors' opinions is determined by the type of trust relationship.

Specifically, in social interactions, individual behaviors are driven by psychological factors, where trust relationships and consensus advantage exert a synergistic effect on opinion propagation. A user is more inclined to accept congruent opinions if their opinion aligns more closely with that of trusted neighbors; conversely, a greater opinion divergence from untrusted neighbors leads to a stronger tendency to reinforce opposing opinions. The corresponding mathematical formulations are as follows: At time t, the opinion similarity c_{ij}^t between node i and its out-degree neighbor j (i.e., the degree of consensus) is defined by Eq (2.2). The opinion update rule for node i from time t to $(t+1)$ is presented in Eq (2.3). Notably, in online social networks, the direction of trust relationships is opposite to that of opinion propagation. Thus, node i is influenced by its out-degree neighbors while simultaneously exerting influence on its in-degree neighbors. It can be seen intuitively from Eq (2.3) that, in mathematical form, the COC model inherits the F-J model, and the normalized weights ensure that the range of opinions always remains within $[-1, 1]$. It should be noted that the continuous opinions in this paper refer to opinions whose values lie in a continuous interval rather than discrete values.

$$c_{ij}^{t}=\begin{cases}1-\dfrac{1}{2}\times\left|o_{j}^{t}-o_{i}^{t}\right|, e_{ij}>0,\\ \dfrac{1}{2}\times\left|o_{j}^{t}-o_{i}^{t}\right|, e_{ij}<0,\end{cases} \tag{2.2}$$

$$o_{i}^{t+1}=\begin{cases}o_{i}^{t}, & \text{if } q=1,\\ \dfrac{1}{1+\sum\limits_{j\in N_{i}^{out}}\omega_{ij}\times c_{ij}^{t}}\times o_{i}^{t}+\sum\limits_{j\in N_{i}^{out}}\left[\dfrac{\omega_{ij}\times c_{ij}^{t}}{1+\sum\limits_{j\in N_{i}^{out}}\omega_{ij}\times c_{ij}^{t}}\times\left(e_{ij}\times o_{j}^{t}\right)\right], & \text{if } q=0,\end{cases} \tag{2.3}$$

where e_{ij} and ω_{ij} denote the trust relationship sign and trust degree from node i to its out-degree neighbor j, respectively; and o_i^t and o_j^t represent the opinions of nodes i and j at time t, respectively. The value of q is in the set $\{0,1\}$, taking the value 0 with an influence probability σ and 1 with a stubbornness probability $(1-\sigma)$. N_i^{out} denotes the set consisting of the out-degree neighbors of node i.

2.2.2. Tie-breaking rule for competing opinions

This paper first defines positive/negative out-degree neighbors. The positive out-degree neighbors of node i are the out-degree neighbor nodes that either hold a positive opinion connected by a positive edge or hold a negative opinion connected by a negative edge. The set of positive out-degree neighbors of node i is denoted as A_i^{out}, as presented in Eq (2.4). Similarly, the negative out-degree neighbors of node i are out-degree neighbor nodes that either hold a negative opinion connected by a positive edge or hold a positive opinion connected by a negative edge. The set of negative out-degree neighbors of node i is denoted as B_i^{out}, as shown in Eq (2.5). Analogously, the set of positive/negative in-degree neighbors of node i can be respectively denoted by A_i^{in} and B_i^{in}. Thus, the set of in-degree/out-degree neighbors of node i is respectively denoted by N_i^{in} and N_i^{out}, as presented in Eqs (2.6) and (2.7), and illustrated in Figure 2.

$$A_{i}^{out}=\left\{v_{j}|o_{j}\in(0,1]\wedge e_{ij}=1\right\}\cup\left\{v_{j}|o_{j}\in[-1,0)\wedge e_{ij}=-1\right\}, \tag{2.4}$$

$$B_{i}^{out}=\left\{v_{j}|o_{j}\in(0,1]\wedge e_{ij}=-1\right\}\cup\left\{v_{j}|o_{j}\in[-1,0)\wedge e_{ij}=1\right\}, \tag{2.5}$$

$$N_{i}^{in}=A_{i}^{in}+B_{i}^{in}=N_{i}^{in+}+N_{i}^{in-}, \tag{2.6}$$

$$N_{i}^{out}=A_{i}^{out}+B_{i}^{out}=N_{i}^{out+}+N_{i}^{out-}, \tag{2.7}$$

where N_i^{in+} denotes the set of in-degree neighbors of node i directly connected by a positive edge, while N_i^{in-} denotes the set of in-degree neighbors of node i directly connected by a negative edge. Similarly, N_i^{in+} denotes the set of out-degree neighbors of node i directly connected by a positive

edge, and N_i^{out-} denotes the set of out-degree neighbors of node i directly connected by a negative edge.

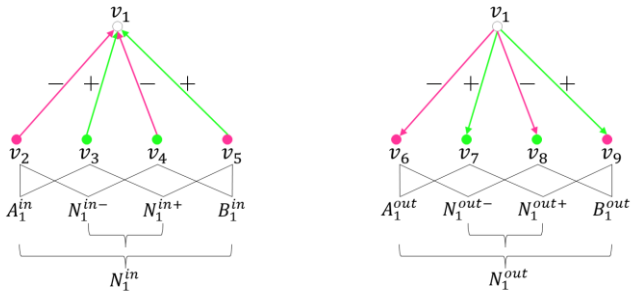

Figure 2. The set of in-degree/out-degree neighbors of node i.

Second, considering the competitive propagation of positive and negative opinions, a competition rule is introduced to optimize the opinion propagation model. If node i has only positive or negative out-degree neighbors, the edge weights of the out-degree neighbors are defined as shown in Eq (2.8) [43], accounting for the average allocation of attention. If node i has both positive and negative out-degree neighbors, the proportional probability tie-breaking rule [9] is incorporated to account for the influence of conformity on opinion adoption [44], where the edge weights of positive/negative out-degree neighbors are defined as shown in Eqs (2.9) and (2.10), respectively.

$$\omega_{ij} = \frac{1}{\left|N_i^{out}\right|}, \tag{2.8}$$

$$\omega_{ij} = \frac{\left|A_i^{out}\right|}{\left|N_i^{out}\right|}, \tag{2.9}$$

$$\omega_{ij} = \frac{\left|B_i^{out}\right|}{\left|N_i^{out}\right|}, \tag{2.10}$$

where ω_{ij} denotes the trust degree of node i in its out-degree neighbor j, and $|\cdot|$ denotes the cardinality of a set.

3. Seeding algorithm

In DQN, the maximization operation in computing the target Q-value may lead to a systematic overestimation bias of the Q-value, which causes instability to the training process. Therefore, this paper employs DDQN [45], leveraging two deep neural networks with the same structure to decouple action selection and reward evaluation, namely the online Q-network and the target Q-network, denoted as $Q_{online}(s,a,b;\theta_{online})$ and $Q_{target}(s,a,b;\theta_{target})$, respectively. The online network is responsible for selecting the optimal action, while the target network is used to evaluate the reward

corresponding to this action. In contrast to the online network, the target network parameters update slowly, thereby enhancing the stability of Q_{target}-value.

DDQN presents a novel solution paradigm, treating the optimal seed selection problem as a parameter estimation task for the reward function (i.e., Q-function). Through learning the optimal seed selection strategy, the agent ultimately generates an optimal seed sequence consisting of k nodes, which maximizes the expected cumulative reward within the total time steps T. This objective is equivalent to maximizing the expected influence coverage in the IM problem. Specifically, at each time step, the current action is first selected following an ε-greedy strategy to obtain an optimal seed. The ε-greedy strategy selects an action randomly with probability ε and chooses the action corresponding to the maximum predicted Q-value with probability $(1-\varepsilon)$, where the predicted Q-value is computed via the Q-network. Subsequently, a MSE loss function is constructed to update the parameters of the Q-network. The MSE loss is formulated based on the TD error, aiming to minimize the difference between the predicted Q-value and the target Q-value, with the Q-network parameters updated using SGD. In the TD error, the target Q-value at the current time step is derived from both the actual reward yielded by the current action and the maximum predicted Q-value of the next state. The discount factor γ effectively balances immediate and delayed rewards, thereby enhancing the accuracy of the expected cumulative reward estimation. It should be noted that the initial value of ε is typically set to 1, followed by a linear decay to 0.05 over 10,000 steps, after which it remains constant. This configuration is designed to balance exploration and exploitation during training. In the early phase, emphasis is placed on exploring new experiences to avoid local optima. In the later phase, priority is given to exploiting accumulated experience to accelerate the learning process.

In this paper, game theory is integrated into the decision-making framework, and the complementary advantages of graph representation learning and DDQN are combined to propose a seeding algorithm (i.e., DM-DDQN) suitable for the TDCIM problem, which endows positive influence with more competitive advantages and strategic initiative. By reconstructing the state space and action space and incorporating the prioritized experience replay (PER) and soft update mechanisms, this algorithm further optimizes the DDQN algorithm, effectively improving the seed quality and training efficiency. Specifically, first, a signed attention mechanism is introduced to enhance the SDGNN algorithm for node embedding learning. Subsequently, the state space of DDQN is redesigned by leveraging the sum of embeddings from positive and negative seeds, thereby enabling fine-grained characterization of the propagation structural features of nodes. Second, actions are simplified into seed selection strategies, and four actions are selected to redesign the action space of DDQN, which significantly reduces the complexity of the action space. Third, in the case where the opponent's strategy is unknown, a mixed-strategy Nash equilibrium framework is introduced for optimal action selection. The four candidate actions are defined as the pure strategy space, and by solving for the mixed-strategy Nash equilibrium, the probability distribution of selecting each candidate action is obtained. The action to be executed is then randomly sampled according to this probability distribution, and finally, the optimal positive seed under the current state is determined, which makes the opponent's payoffs as balanced as possible while keeping the focal player's strategy unpredictable. Finally, the PER mechanism is adopted to update the parameters of the online network, and the soft update mechanism is used to update the parameters of the target network. The advantage of the PER mechanism is that it updates the parameters using only a subset of samples, which eliminates the correlation between samples, not only reducing the computational load but also helping to improve the training efficiency. The soft update mechanism ensures the smooth transition of the

target Q-values, further enhancing the stability of training.

3.1. State space

The state space design in DDQN is required to reflect the overall network state, including the influence coverage of selected positive and negative seeds, as well as the potential influence of candidate seeds. Unlike network state representations that rely on hand-crafted features [46], drawing inspiration from the integration of network embedding with DQN for addressing IM problems [22–24], this paper leverages directed signed graph embedding to design the state space, effectively capturing node positional information within the network structure, thereby facilitating the accurate evaluation of the marginal influence gain of the newly added seed. Specifically, we first introduce a signed attention mechanism [47,48] to optimize the SDGNN algorithm, proposing a directed signed graph embedding algorithm based on signed attention (S-SDGNN), where the calculation of signed attention weights is shown in Eq (3.1). Subsequently, the embeddings of selected positive and negative seeds are summed to form the current state, as illustrated in Eq (3.2). It should be noted that we chose SDGNN as the foundation, as it combines structural balance theory with social status theory, thereby enabling the full capture of social mechanisms within directed signed networks.

$$\lambda_1 = \frac{1}{\left|N_i^{out+}\right|}, \quad \lambda_2 = \frac{1}{\left|N_i^{out-}\right|}, \quad \lambda_3 = \frac{1}{\left|N_i^{in+}\right|}, \quad \lambda_4 = \frac{1}{\left|N_i^{in-}\right|}, \tag{3.1}$$

$$s_t = \sum\nolimits_{v_i \in S_t^+ \cup S_t^-} z_i, \tag{3.2}$$

where λ_1, λ_2, λ_3, and λ_4 denote the signed attention weights corresponding to the out-degree neighbors of v_i connected by a positive edge, out-degree neighbors of v_i connected by a negative edge, in-degree neighbors of v_i connected by a positive edge, and in-degree neighbors of v_i connected by a negative edge, respectively. z_i represents the embedding of v_i, s_t denotes the state at time t, while S_t^+ and S_t^- stand for the sets of selected positive and negative seeds at time t, respectively.

3.2. Action space

In general, an action is defined as the selection of a seed. To compress the action space and mitigate computational costs, drawing upon the action design paradigm for OM problems [13,14], this paper simplifies actions to the selection of a seed choice strategy, reconstructing the action space using four simple yet intuitive strategies: positive expansion, negative blocking, collective influence, and community centrality. The first two strategies respectively focus on the promotion of positive influence and the suppression of negative influence, while the latter two consider the breadth and depth of influence diffusion, effectively resolving the conflict between seed quality and computational efficiency. Detailed explanations of these four strategies are as follows:

(1) Positive expansion: The candidate seed with the maximum number of in-degree neighbors connected by a positive edge is selected as the optimal positive seed, aiming to promote the spread of positive influence.

(2) Negative blocking: The candidate seed with the maximum number of negative out-degree

neighbors is selected as the optimal positive seed, aiming to suppress the diffusion of negative influence.

(3) Collective influence [49]: The candidate seed with the maximum collective influence is selected as the optimal positive seed. In this paper, collective influence with a radius of 2 is employed, which considers the influence of both the node itself and its second-order neighbors, as mathematically expressed in Eq (3.3).

$$CI\left(v_i\right)=\left(\left|N_i\right|-1\right)\times\sum_{dist\left(v_i,v_j\right)=2}\left(\left|N_j\right|-1\right), \tag{3.3}$$

where v_j denotes the second-order neighbor of v_i, meaning the shortest path distance between the two nodes is 2.

(4) Community centrality [50]: The candidate seed with the maximum internal and external community sizes is selected as the optimal positive seed. Its mathematical formulation is given in Eq (3.4).

$$CbC\left(v_i\right)=\frac{1}{N}\sum_{\left(v_j\in N_i\right)\wedge\left(C_i=C_j\right)}\left|C_j\right|+\frac{1}{N}\sum_{\left(v_j\in N_i\right)\wedge\left(C_i\neq C_j\right)}\left|C_j\right|, \tag{3.4}$$

where C_i and C_j denote the communities to which nodes i and j belong, respectively, and $|\cdot|$ represents the number of nodes within a community.

Second, inspired by the optimal strategy solution framework for the OM problem [12], this paper introduces a mixed-strategy Nash equilibrium framework to analyze the TDCIM problem. The Nash equilibrium mixed strategy of the focal player is regarded as the optimal rational choice, which guides the agent to select the optimal action when the opponent's strategy is unknown. A mixed strategy is a probability distribution over a set of pure strategies, from which a player randomly selects a pure strategy. A pure strategy can be regarded as a special case of mixed strategy, where the selection probability of one pure strategy is 1, and the selection probabilities of all remaining strategies are 0. Based on the definition of the TDCIM problem and its reward function, our payoff is the number of net positive nodes, while the opponent's payoff is the number of net negative nodes, indicating that the payoffs of the two sides are completely opposed and are negatives of each other; therefore, this constitutes a two-player zero-sum game. In zero-sum games, when no pure-strategy Nash equilibrium exists, if the focal player adopts a fixed pure strategy, it is easily countered by the opponent in a targeted manner. Therefore, under conditions of unknown opponent's strategy, a mixed strategy is the optimal rational choice for the focal player. This is also well understood in the real world: Almost any deterministic strategy will have a counter strategy against it, and no strategy can always gain the upper hand, so the optimal strategy usually has a certain randomness.

A mixed-strategy Nash equilibrium serves as the optimal and stable solution under unpredictable conditions, achieving a high degree of integration between the stability of the opponent's payoff and the randomness of the focal player's strategy. It ensures that the opponent's payoff is distributed as evenly as possible while rendering the focal player's strategy unpredictable. This paper reasonably assumes that both parties seek to maximize their own benefits. When the opponent's strategy is unpredictable, the focal player's optimal response is to equalize the opponent's expected payoff across all possible choices. If payoffs are unevenly distributed among different options, the opponent will inevitably select the strategy with a higher payoff. Nash [51] demonstrated that for non-cooperative games with finite strategy sets, pure-strategy Nash equilibria may not exist, yet mixed-strategy Nash

equilibria are guaranteed to exist. In other words, the mixed-strategy Nash equilibrium constitutes a more generalized form of equilibrium. For ease of understanding, this paper uses a strategy space containing two strategies as an example and applies the payoff-equating method to solve for our side's Nash equilibrium mixed strategy (hereinafter referred to as the equilibrium strategy). Assume that a fully supported mixed-strategy Nash equilibrium exists, meaning that all pure strategies of both sides are played with positive probability in the equilibrium strategy. The payoff matrix for the two players across all strategy combinations is presented in Table 3, where Player 1 denotes the focal player, and Player 2 denotes the opponent.

Table 3. Payoff matrix.

1	2	
	u_1'	u_2'
u_1	$Q(s,u_1,u_1'),\ -Q(s,u_1,u_1')$	$Q(s,u_1,u_2'),\ -Q(s,u_1,u_2')$
u_2	$Q(s,u_2,u_1'),\ -Q(s,u_2,u_1')$	$Q(s,u_2,u_2'),\ -Q(s,u_2,u_2')$

According to the payoff-equating method, where the opponent's expected payoffs across different strategies are assumed to be equal, we formulate the equation as presented in Eq (3.5). Solving this equation yields the probability ρ. Thus, for the current state s, the focal player's Nash equilibrium mixed strategy consists of selecting strategy u_1 with probability ρ and strategy u_2 with probability $(1-\rho)$.

$$\rho \times Q(s,u_1,u_1') + (1-\rho) \times Q(s,u_2,u_1') = \rho \times Q(s,u_1,u_2') + (1-\rho) \times Q(s,u_2,u_2'), \tag{3.5}$$

where u_1 and u_2 denote our two pure strategies, u_1' and u_2' denote the opponent's two pure strategies, ρ represents the probability of the focal player selecting u_1, and $(1-\rho)$ represents the probability of the focal player selecting u_2.

For a strategy space containing four strategies, the primal linear programming problem is formulated according to the minimax theorem, as shown in Eq (3.6). This paper first solves the primal linear programming problem using the simplex method to obtain our side's equilibrium strategy, and then randomly selects one of the four actions according to this probability distribution to obtain the optimal seed selection strategy for the current state.

$$\begin{aligned} &\max\nolimits_{\rho_1,\rho_2,\rho_3,\rho_4,\lambda}\ \lambda \\ &\text{s.t.}\quad \sum_{i=1}^{4} \rho_i \times Q_{ij} \geq \lambda,\ j = 1,2,3,4, \\ &\sum_{i=1}^{4} \rho_i = 1, \\ &\rho_i \geq 0,\ i = 1,2,3,4, \end{aligned} \tag{3.6}$$

where ρ_i denotes the probability assigned to our i-th pure strategy, Q denotes our payoff matrix, and λ denotes our equilibrium payoff.

3.3. Reward function and training process

(1) Reward function

The reward function is defined as the propagation increment of dominant competitive influence resulting from the addition of a new node to the selected seed set at time t, and its mathematical expression is given in Eq (3.7).

$$\Gamma_t\left(v^*\right)=\left[I_t\left(S_t^+\cup\left\{v^*\right\}\right)-I_t\left(S_t^-\cup\left\{v^*\right\}\right)\right]-\left[I_t\left(S_t^+\right)-I_t\left(S_t^-\right)\right], \tag{3.7}$$

where v^* denotes the newly added node, and S_t^+ and S_t^- represent the positive seed set and the negative seed set at time t, respectively. $I_t(\cdot)$ is the propagation function, which describes the influence scope of the seed set after propagation at time t.

(2) Online network training

This paper adopts the PER mechanism [52] for the training of the online network. Specifically, at each time step, the online network first selects the optimal action, generates a new experience, and stores it in the replay buffer. The initial priority of the new experience is set to 1, while the priorities of existing experiences are updated based on their TD errors. Next, a batch of experiences is sampled from the replay buffer according to their sampling priority probabilities. A MSE loss function is then constructed using the TD errors of this batch of experiences. Finally, SGD is employed to minimize the loss function and update the parameters of the online network. Instead of directly utilizing the latest samples or uniformly sampling from existing ones, this mechanism prioritizes the learning of high-value samples. On the basis of fully leveraging existing samples, it further enhances the utilization efficiency of high-value samples (i.e., experiences with large absolute TD errors) and accelerates the convergence rate of model training.

The TD error δ_i for experience i is shown in Eq (3.8).

$$\delta_i=r_i+\gamma\times Q_{target}\left(s_i',\underset{a_i'}{\operatorname{argmax}}\,Q_{online}\left(s_i',a_i',b_i'\right),b_i'\right)-Q_{online}\left(s_i,a_i,b_i\right), \tag{3.8}$$

where experience $i=(s_i,a_i,b_i,r_i,s_i')$, s_i denotes the current state (i.e., the state at the current time step), and s_i' denotes the next state (i.e., the state at the next time step); a_i and a_i' represent the agent's current action and next action, respectively; b_i and b_i' represent the opponent's current action and next action, respectively; r_i indicates the actual reward obtained at the current time step (i.e., the reward yielded by executing the current action), which is derived through simulation under the COC model. $Q_{target}\left(s_i',\underset{a_i'}{\operatorname{argmax}}\,Q_{online}(s_i',a_i',b_i'),b_i'\right)$ indicates first selecting the action corresponding to the maximum expected Q-value at the next time step using the online network, then computing the expected Q-value of this selected action using the target network. $Q_{online}(s_i,a_i,b_i)$ represents the predicted Q-value at the current time step calculated by the online network.

The priority p_i of experience i is shown in Eq (3.9).

$$p_i=\left(\left|\delta_i\right|+\xi\right)^{\alpha}, \tag{3.9}$$

where ξ is a small constant introduced to avoid a zero priority, and it is set to 10^{-6} in this paper. α denotes the priority exponent, which is used to adjust the priority, and its value is set to 0.6 in this paper.

The sampling priority probability P_i of experience i is shown in Eq (3.10).

$$P_i = \frac{p_i}{\sum_{j=1}^{\mu_2} p_j}, \tag{3.10}$$

where μ_2 denotes the total number of experiences in the replay buffer.

The weight w_i corresponding to experience i is shown in Eq (3.11); the adjusted loss function is shown in Eq (3.12).

$$w_i = \frac{\left(\frac{1}{\mu_1} \times \frac{1}{p_i}\right)^{\beta}}{\max_{p_i}\left(\frac{1}{\mu_1} \times \frac{1}{p_i}\right)^{\beta}}, \tag{3.11}$$

$$loss = \frac{1}{\mu_3} \times \sum_{i=1}^{\mu_3} w_i \times \delta_i^2, \tag{3.12}$$

where μ_1 denotes the size of the replay buffer, μ_3 denotes the sampling batch size, and β denotes the bias correction coefficient for priority sampling. Specifically, β is typically initialized to 0.4, incremented by 10^{-4} at each step, and capped at a maximum value of 1.

(3) Target network training

In this paper, a soft update mechanism [45] is adopted to train the target network, as formulated in Eq (3.13). Specifically, this mechanism performs a smooth update on the parameters of the target network using those of the online network, thereby mitigating fluctuations in the target Q-values and improving the stability of the training process.

$$\theta'_{target} = \tau \times \theta'_{online} + (1-\tau) \times \theta_{target}, \tag{3.13}$$

where θ_{target} denotes the parameters of the target network at the current time step, whereas θ'_{online} and θ'_{target} represent the parameters of the online network and the target network at the next time step, respectively. τ denotes the soft update coefficient.

In summary, the pseudocode of the DM-DDQN algorithm proposed in this paper is presented in Algorithm 1. Its main steps are summarized as follows: (a) Data initialization: Initialize the seed set as an empty set, compute the embedding representation for each node, and derive the initial state s_1 (Steps 1–2). (b) Seed selection: Leverage an ε-greedy strategy combined with a mixed-strategy Nash equilibrium; actions are selected randomly with probability ε, while the focal player's Nash equilibrium mixed strategy is chosen with probability $(1-\varepsilon)$, thereby obtaining the optimal positive seed under the current state s_i (Steps 4–14). (c) Parameter update: First, compute the actual reward r_i and the next state s_{i+1}, and add the new experience to the replay buffer. Next, sample a batch of experiences via priority sampling and construct the MSE loss based on the TD error. Finally, update the parameters θ_{online} of the online network using SGD, and perform a soft update on the parameters θ_{target} of the target network using θ_{online} (Steps 15–18). (d) Repeat steps (b) and (c) until k positive

seeds are selected, yielding the optimal positive seed set.

Algorithm 1. The DM-DDQN algorithm

Input: Directed signed social network $G(V,E)$, the number of positive seed nodes k, the size of the replay buffer μ_1, the batch size of priority sampling μ_3;

Output: Optimal positive seed set S^+;

Initialize $S^+ = \emptyset,\ \varepsilon = 1$;
Calculate the embeddings of each node in V and the initial state s_1;
for $i = 1$ to k, **do**
    Select the action a_i using ε-greedy strategy and mixed-strategy Nash equilibrium, where the payoff matrix is calculated using the online network $Q_{online}(s_i, a, b; \theta_{online})$, and s_i denotes the current state;
    if $a_i =$ Positive Expansion, **then**
        Select the candidate seed with the maximum number of in-degree neighbors connected by a positive edge as the optimal positive seed v^*;
    else if $a_i =$ Negative Blocking, **then**
        Select the candidate seed with the maximum number of negative out-degree neighbors as the optimal positive seed v^*.
    else if $a_i =$ Collective Influence, **then**
        Select the candidate seed with the maximum collective Influence with a radius of 2 as the optimal positive seed v^*.
    else if $a_i =$ Community Centrality, **then**
        Select the candidate seed with the maximum internal and external community sizes as the optimal positive seed v^*.
    end if
    $S^+ = S^+ \cup \{v^*\}$;
    Calculate the actual reward $r_i = \Gamma_i(v^*)$ and the next state s_{i+1};
    Add the experience $(s_i, a_i, b_i, r_i, s_{i+1})$ to the replay buffer;
    update the online network parameters θ_{online} using priority sampling， MSE and SGD;
    Soft update the target network parameters θ_{target} with θ_{online};
    update $\varepsilon = max(0.05, \varepsilon - 0.000095)$;
end for

The time complexity of Algorithm 1 is primarily governed by two components: the preprocessing stage for network embedding and the online inference stage of DDQN, with their respective time complexities being $O(L \times |E| \times d)$ and $O(k \times H \times d)$. Thus, the overall time complexity of the seed selection process is $O(L \times |E| \times d + k \times H \times d)$, where L denotes the number of stacked layers in SDGNN, d represents the node embedding dimension, $|E|$ stands for the total number of edges, H indicates the hidden layer size of the Q-network, and k is the number of positive seeds.

4. Experiment and analysis

This paper proposes a truth source selection strategy tailored for directed signed social networks. First, by incorporating the echo chamber effect and structural balance theory and integrating the

complementary advantages of the CIC model and F-J model, we propose a continuous opinion competition model (i.e., COC). This model comprehensively characterizes the processes of opinion activation and updating, thereby reflecting the laws of rumor propagation in the real world. Second, in the scenario where the opponent's strategy is unknown, we introduce a game-theoretic analytical framework. By leveraging directed signed graph embedding and mixed-strategy Nash equilibrium, we redesign the state space and action space of DDQN and propose an optimal positive seed set selection strategy (i.e., DM-DDQN). This strategy endows truth with a greater competitive advantage, thereby effectively suppressing rumor propagation.

To validate the effectiveness of the proposed algorithm, four real-world online social network datasets are selected, namely Bitcoin-Alpha [53], Bitcoin-OTC [53], Wiki-RfA [54], and Slashdot [55]. Their statistical characteristics are presented in Table 4. All four datasets belong to directed signed social networks, yet they vary in scale, average degree, and proportion of positive edges, which facilitates a comprehensive evaluation of the seeding algorithm's generalization ability. Specifically, Bitcoin-Alpha and Bitcoin-OTC model users of Bitcoin trading platforms as nodes, where the signs of edges correspond to positive or negative trust ratings; Wiki-RfA models Wikipedia members as nodes, with edge signs denoting votes of support or opposition; Slashdot models news comment users as nodes, where edge signs represent friend or enemy labels.

Table 4. Statistical information of real datasets.

Networks	Nodes	Edges	Average degree	Positive edges (%)
Bitcoin-Alpha	3783	24186	12.79	93.65
Bitcoin-OTC	5881	35592	6.05	89.99
Wiki-RfA	10,835	159,388	14.71	77.95
Slashdot	82,144	549,202	6.68	77.40

All experiments were conducted on Windows 11 with the following hardware configuration: an Intel(R) Core(TM) i9-14900HX CPU, an NVIDIA GeForce RTX 4090 24 GB GPU, 32 GB of RAM, and a 1 TB hard drive. All implementations were done using Python 3.11 and PyTorch 2.2. Key hyperparameters were configured as follows: The replay buffer size was set to 10,000, and the batch size was 32. For DDQN, both the online network and target network were designed as fully connected feedforward neural networks with two hidden layers, each containing 256 neurons, and ReLU was adopted as the activation function. The node embedding dimension was set to 64, the influence probability σ to 0.2 (it can also be uniformly selected from {0.2, 0.05, 0.01}; see Section 4.3.1), the discount factor γ to 0.95, the soft update coefficient τ to 0.005, and the initial learning rate φ to 0.001. The Adam optimizer was employed for model training.

All experiments were independently repeated 20 times, and the average values were reported to ensure the reliability of the results. Bootstrap confidence intervals (CIs) at a 95% confidence level were used to characterize the uncertainty of the mean. In addition, the repeated experiments included independent model training, opponent selection, and propagation simulation. This paper assumes that the number of positive and negative seeds is equal and that both sides share the same strategy space; positive seeds are selected first, followed by negative seeds. Notably, to align with the TDCIM task, the performance metric for seeding algorithms in this paper is defined as the difference between the number of positive opinions and negative opinions in the network after a given propagation time T, which captures the extent to which the influence range of the positive seed set exceeds that of the negative seed set.

4.1. Comparative experiment

To validate the superiority of the DM-DDQN algorithm proposed in this paper, seven representative seeding algorithms are selected as comparison benchmarks, namely integrated strategy, FICLF [19], K-step greedy [18], S-CMIA [20], DGN [24], ToupleGDD [23], and QOMF [14]. To ensure the comparability of algorithm performance, the propagation models of the compared algorithms were appropriately replaced, retaining only the proposed selection strategy and a subset of the coupled propagation rules. The seed sets were then regenerated under the same COC model. All baselines use the same seed budget and adversary strategy as this paper and are evaluated under the same experimental environment and computational resources. For comparative algorithms that do not consider signed edges, competitors, or limited time, the above information can be directly ignored; for example, all negative edges can be treated as positive edges. For the other parameters in the baseline algorithms, we retain the recommended settings from their original papers. The experimental design is as follows: First, we assume that the adversary adopts a random strategy (i.e., randomly selecting one of the four strategies with equal probability), fix the propagation duration, vary the number of seed nodes, and compare the positive seed quality of all algorithms. Second, we fix the number of seed nodes, vary the propagation duration, and compare the positive seed quality of all algorithms. Finally, when the adversary switches to the equilibrium strategy (obtained using the proposed algorithm), we fix the propagation duration, vary the number of seed nodes, and compare the positive seed quality of all algorithms.

(1) Integrated strategy: This strategy simultaneously considers the positive expansion and negative blocking strategies, selecting the candidate seed with the maximum number of in-degree neighbors connected by a positive edge and negative out-degree neighbors connected as the optimal seed.

(2) FICLF: A finite-time positive influence maximization algorithm for directed signed networks, which only considers the positive activation of positive seeds through a positive edge.

(3) K-step greedy: Another finite-time positive influence maximization algorithm for directed signed networks, which accounts for both positive activation via a positive edge and negative activation via a negative edge by positive seeds.

(4) S-CMIA: An algorithm for maximizing the difference between the final positive and negative influences in directed signed networks, which takes into account the competitive propagation scenario of positive and negative information.

(5) DGN: A learning framework integrating GNN and DQN, applicable to large-scale social networks.

(6) ToupleGDD: A learning framework integrating GNN and DDQN, applicable to social networks.

(7) QOMF: A Q-learning framework for COM problems.

In this paper, the time step is fixed at 50, and the number of seeds is varied from 0 to 50. Experiments are conducted on four real-world datasets to compare the propagation performance of different algorithms, as shown in Figure 3. It should be noted that in this paper, curves represent the means, and the shaded regions represent the 95% bootstrap CIs. For brevity, only the upper bounds of the CIs for all baselines are shown. Meanwhile, during the testing (running) phase, convergence is assessed for each simulation sample, and the mean and confidence interval of the strategy payoff are computed only for converged samples, thereby excluding spurious payoffs arising from phase shifts in periodic oscillations to ensure a fair comparison of algorithm performance. A sample is deemed converged if the infinity norm of the difference between the overall opinion vectors at two consecutive steps is less than 10^{-4}. If the propagation continues up to the maximum time step of 2000 without

satisfying the above convergence condition, the sample is deemed to be oscillating, meaning that no steady state exists in the overall opinion distribution. This paper statistically analyzes the convergence rates of different strategies, and the convergence rate of the equilibrium strategy is slightly higher than that of the random strategy. On the four datasets, the convergence rates of the equilibrium strategy are above 0.95, 0.95, 0.85, and 0.7. Since the topology of the Slashdot dataset itself tends to induce oscillations, the convergence rate of the equilibrium strategy is relatively low; however, the influence evaluation on the converged samples remains valid.

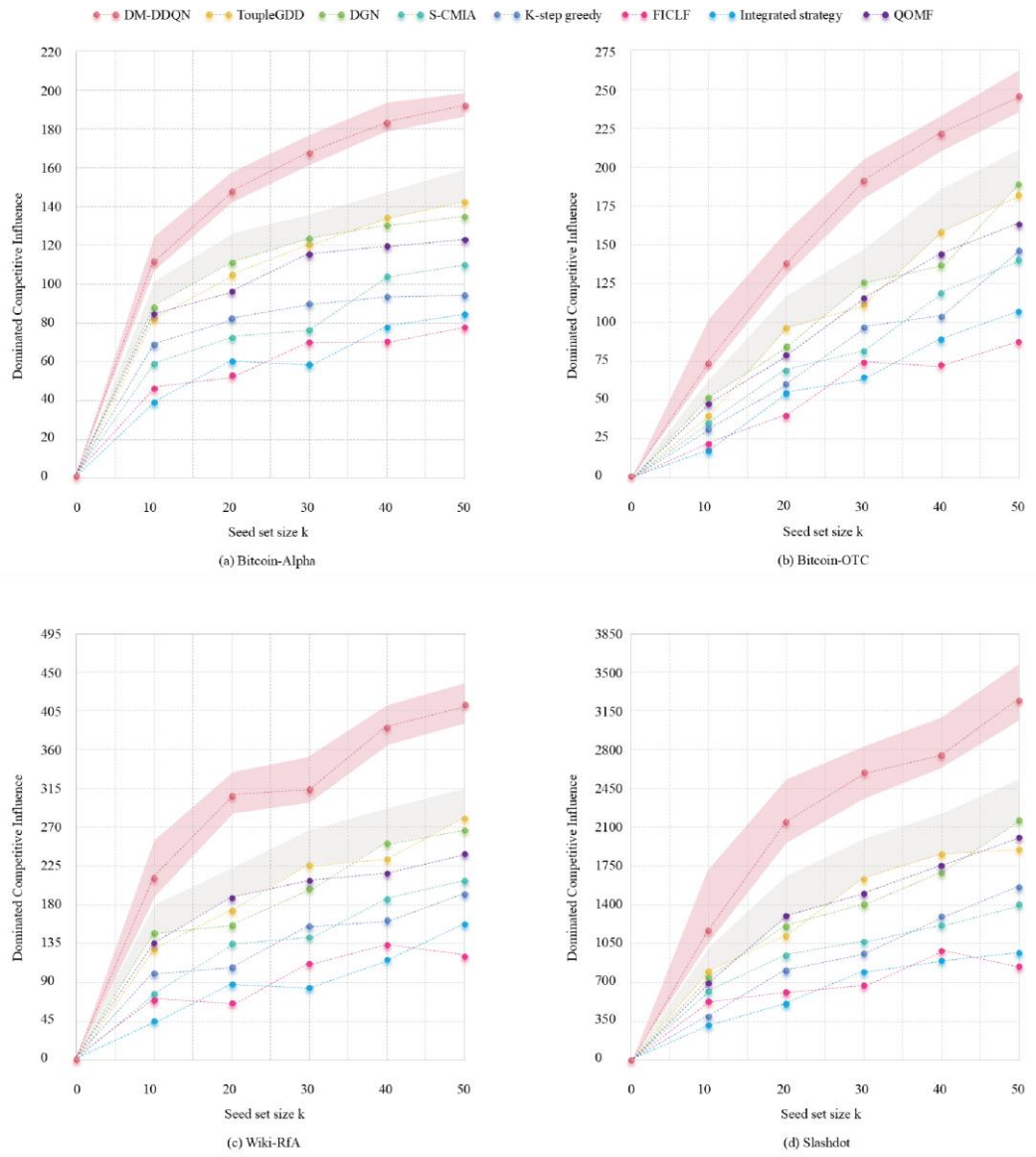

Figure 3. Performance comparison of seeding algorithms across different seed counts. (T = 50 and the opponent uses a random strategy.)

The above experimental results are analyzed as follows: First, the proposed DM-DDQN algorithm outperforms all baseline algorithms across the four datasets, indicating its strong generalization capability across networks with varying scales and topologies. The selected positive seed sets

demonstrate superior influence propagation capacity and a clear competitive advantage over negative seed sets. This is because the DM-DDQN algorithm is specifically tailored to address the TDCIM problem. Compared with integrated strategy, our approach adopts a mixed-strategy Nash equilibrium, enabling optimal rational decision-making under an uncertain opponent's strategy. This approach not only equalizes the opponent's potential payoffs as much as possible but also preserves the stochasticity of the focal player's strategy. In contrast to the FICLF algorithm, we integrate structural balance theory to account for mixed paths consisting of both positive and negative edges. Unlike the K-step greedy algorithm, our method addresses the scenario with an unknown opponent's strategy, which better aligns with real-world rumor control dynamics in online social networks. Compared with the S-CMIA algorithm, our algorithm incorporates temporal factors into the decision-making framework, balancing immediate and long-term payoffs. Relative to the ToupleGDD and DGN algorithms, this paper models the echo chamber effect and uses game theory to optimize strategy selection, ultimately obtaining the DM-DDQN algorithm. This makes the agent's decision-making process more consistent with the characteristics of human communication behavior, so the selected positive seed set has stronger competitive propagation capability and better interpretability. Compared with the QOMF algorithm, this paper reconstructs the state space and action space of DDQN, making it more suitable for competitive diffusion scenarios in signed social networks.

Second, as the number of seed nodes increases, the advantage of the DM-DDQN algorithm becomes increasingly pronounced. This is attributed to its proactive deployment of strategically critical nodes in the dissemination process, highlighting the pivotal role of seizing the public opinion high ground in competitive diffusion. Such a strategy effectively suppresses opponents, rapidly gains situational control, and thereby secures a dominant position in the competition. Although the marginal gain from increasing positive seeds exhibits fluctuations on the Wiki-RfA dataset, this phenomenon can be attributed to the high proportion of negative edges in the network. While positive seeds activate positive opinions, they concurrently trigger negative opinions via negative edges, resulting in a slight decline in the growth rate of propagation performance. However, compared with other seeding algorithms, our proposed algorithm still maintains a high level of performance, which indicates its excellent robustness and stability.

In this paper, the number of seeds is fixed at 50, while the time step is varied from 0 to 50. Experiments are conducted on four real-world datasets to compare the performance of different seeding algorithms, as shown in Figure 4.

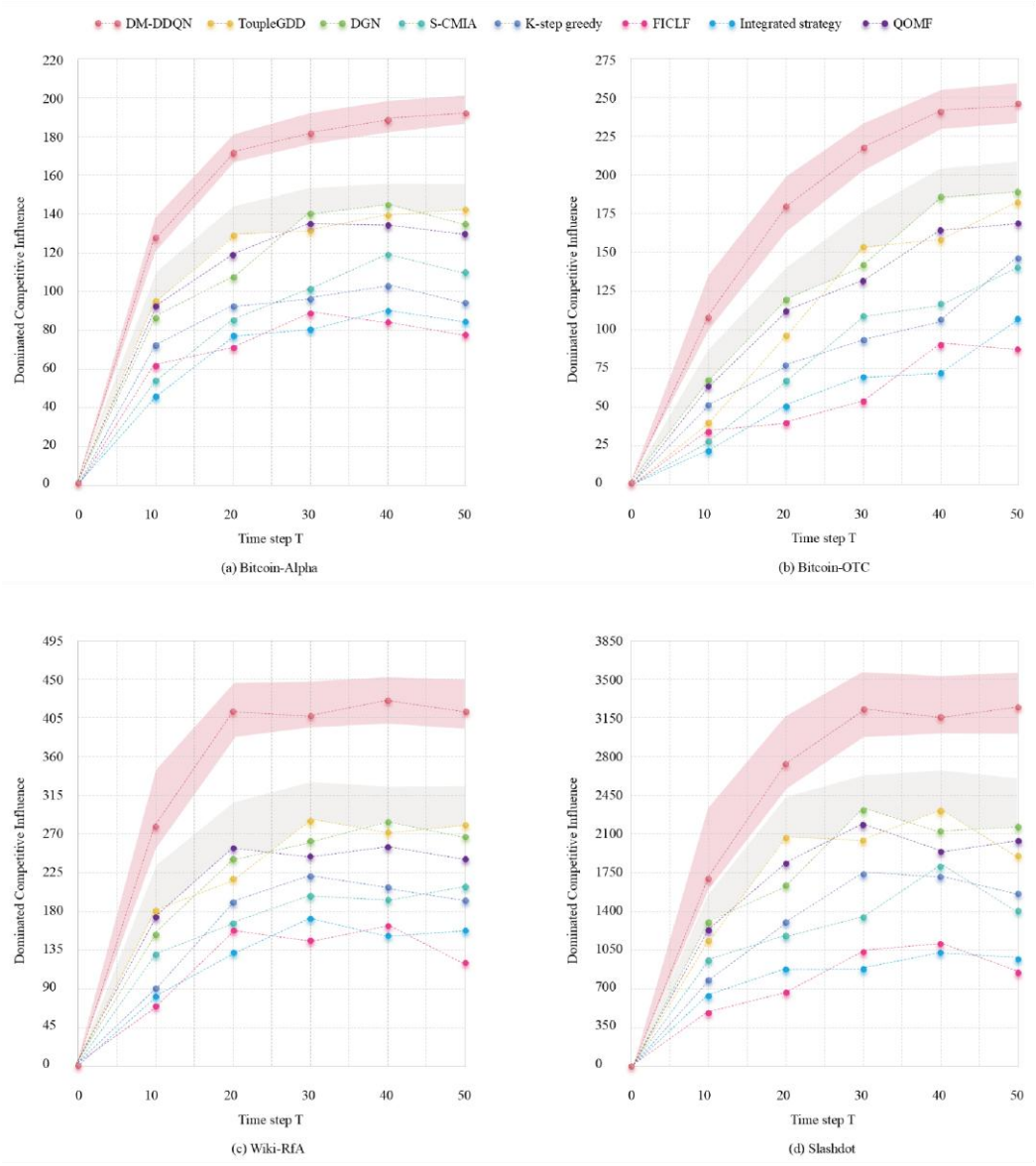

Figure 4. Performance comparison of seeding algorithms across different time steps. (k = 50 and the opponent adopts a random strategy.)

The above experimental results are analyzed as follows: First, across four real-world datasets with distinct scales and structural characteristics, the proposed algorithm consistently outperforms other baseline methods in propagation performance. Furthermore, its competitive advantage remains relatively stable as the number of time steps increases. For the Slashdot dataset, which is characterized by a large total number of nodes and a low average degree, although positive seeds require a certain number of time steps to propagate their influence on the remote regions of the network, the DM-DDQN algorithm still rapidly establishes a dominant position.

Second, as the propagation duration increases, the overall performance of all seeding algorithms initially exhibits a significant upward trend, followed by a gradual deceleration in growth and even a tendency toward decline. This indicates that when the propagation duration exceeds a specific threshold, the dominant competitive influence of the positive seed set may diminish. Therefore, it is

imperative to conduct a quantitative analysis of truth source deployment schemes that account for the number of deployments and their duration, in order to identify the optimal cost-benefit ratio, achieve the intended effect of suppressing rumor propagation, and enhance the utilization efficiency of governance resources. This paper constructs a three-dimensional matrix of maximum expected returns based on the number of seeds and propagation duration, which characterizes the maximum competitive advantage achievable under different resource combinations. This matrix facilitates flexible adjustment of truth source deployment strategies by decision-makers, enabling efficient allocation of debunking resources and maximizing the effectiveness of rumor governance. By integrating Figures 3 and 4 to generate Figure 5, we derive the maximum expected returns under various budget constraints by fixing either the number of time steps or the number of seeds. Specifically, the range is determined based on the constraints on k and T, and then a search algorithm is used to find the strategy that maximizes the payoff. Notably, since distrust relationships reverse the polarity of influence, the optimal truth source deployment strategy is often closely related to the specific network structure and scale.

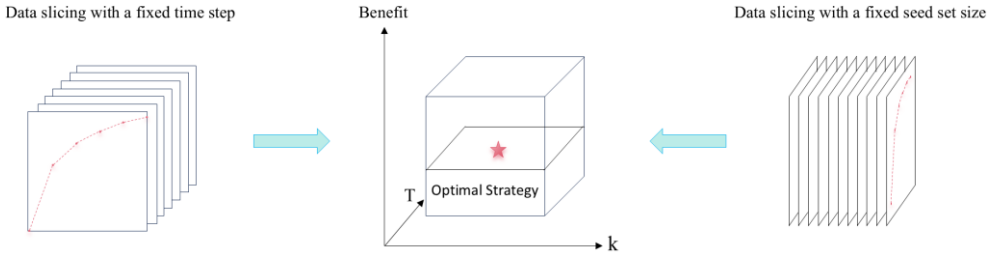

Figure 5. Schematic diagram of maximum dominant competitive influence under varying positive seed deployment strategies.

When the opponent's strategy is changed to an equilibrium strategy, the time step is fixed at 50, and the number of seeds varies from 0 to 50. Experiments are conducted on four real-world datasets to compare the propagation performance of different algorithms, as shown in Figure 6. It should be noted that, since the equilibrium strategies of both parties are obtained using the algorithm proposed in this paper, to avoid comparison bias caused by both parties adopting the same strategy, the two DM-DDQNs are trained separately and independently during the training phase. During the testing phase, the weights of both DM-DDQNs are fully fixed, and no parameter updates or random exploration is performed, so as to ensure the fairness of the comparison of our strategies.

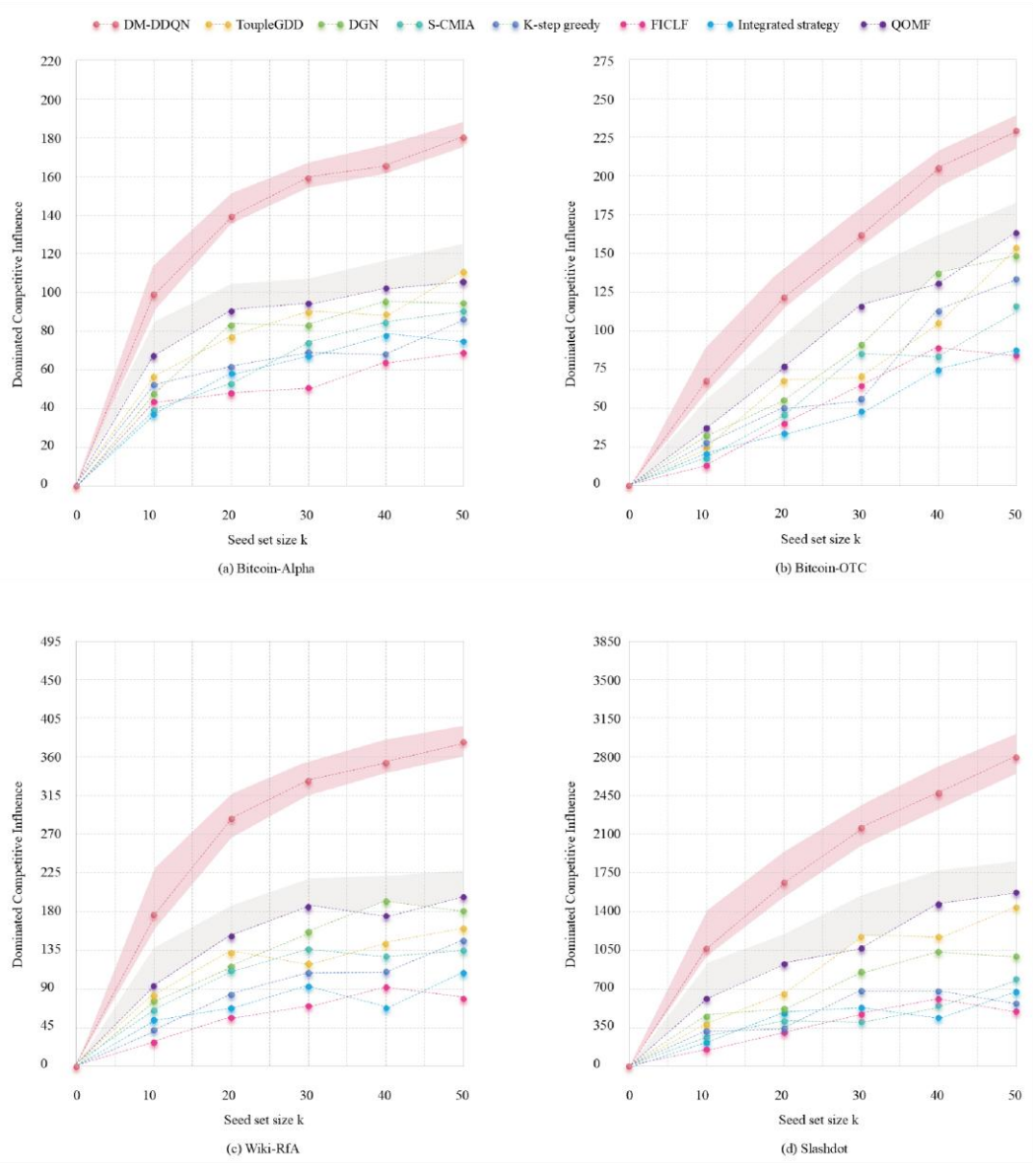

Figure 6. Performance comparison of seeding algorithms across different seed counts. (T = 50 and the opponent adopts an equilibrium strategy.)

As can be seen from Figure 6, when the opponent switches from a random strategy to an equilibrium strategy, the performance of all seeding algorithms declines. However, the algorithm proposed in this paper still maintains the greatest competitive advantage even in the worst case. This is because, when the opponent's strategy is unknown, choosing an equilibrium strategy on our side can, on the one hand, make the opponent's payoff as balanced as possible, and on the other hand, ensure a guaranteed payoff when facing an advanced opponent. In other words, it maximizes our minimum payoff, which is exactly what the minimax theorem states.

4.2. Ablation study

The solution to the traditional IM problem comprises two core components: an information propagation model and an influence maximization algorithm. Accordingly, our proposed solution for

the TDCIM problem primarily consists of the COC model and the DM-DDQN algorithm. To validate the effectiveness of these two modules, this paper designs four variants for ablation experiments to quantify their contributions to dominant competitive influence propagation. Specifically, the CIC model and DDQN algorithm are first adopted as baseline methods. Subsequently, based on the COC model and DDQN algorithm, the state space and action space of DDQN are redesigned, ultimately yielding the DM-DDQN algorithm. The number of seeds and propagation duration are fixed at 50, and experiments are conducted on four real-world online social networks to analyze the contribution of each component to the overall performance, as illustrated in Figure 7. It should be noted that to eliminate dimensional discrepancies across datasets of varying scales, all experimental results in this section are normalized. The evaluation metric employed herein is the ratio of the difference between positive and negative influences, specifically defined as the ratio of the difference between the number of positive and negative opinions in the network to the total number of nodes in the network.

(1) Variant a: CIC model combined with DDQN.

(2) Variant b: COC model combined with DDQN.

(3) Variant c: COC model combined with DDQN incorporating a reconstructed state space.

(4) Variant d: COC model combined with DDQN incorporating a reconstructed state space and action space.

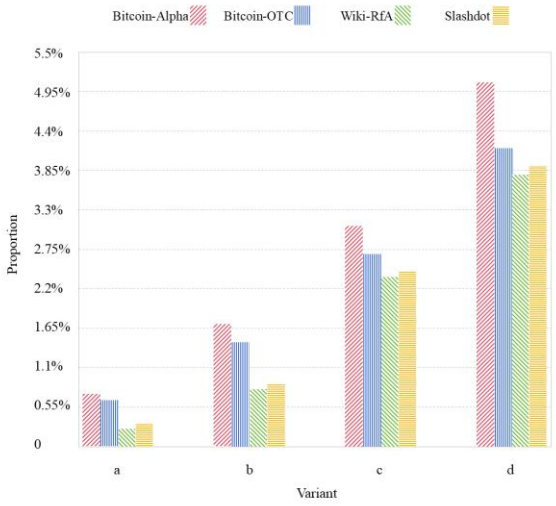

Figure 7. Comparison of propagation performance across different variants. Among them, a) serves as the baseline, b) improves the propagation rules, c) redesigns the state space of DDQN, and d) redesigns the action space of DDQN.

The experimental results presented above indicate that, with variant a) serving as the baseline, all other variants contribute to the propagation performance of the DM-DDQN algorithm, though the contribution of each module exhibits variability across the four datasets. This sufficiently demonstrates that the components of the proposed algorithm synergistically collaborate to maximize dominant competitive influence, thereby achieving state-of-the-art propagation performance among peer methods. Specifically, for variant b), optimizing the propagation model yields a significant improvement in the performance of the seeding algorithm. This improvement can be attributed to two key enhancements: first, the integration of social psychology theories to refine the rules governing opinion propagation and competition; second, the combination of the CIC model and F-J model to

explicitly account for the processes of opinion activation and updating, which enables a fine-grained characterization of the dynamic evolution of continuous opinions and thereby enhances the agent's sensitivity to subtle shifts in the propagation landscape. For variant c), reconstructing the state space to improve the DDQN algorithm leads to a substantial enhancement in the propagation performance of the proposed algorithm. This improvement originates from the use of directed signed graph embeddings to characterize network states. By capturing the higher-order structural features of nodes within the network, this approach enables a more accurate evaluation of the potential propagation capacity of candidate seeds, thus boosting the agent's predictive accuracy with respect to environmental changes. Finally, for variant d), reconstructing the action space to refine the DDQN algorithm further enhances the competitive propagation capability of the positive seed set. This improvement can be attributed to two key design choices: first, the consideration of the scenario with unknown opponent's strategy; second, the adoption of a mixed-strategy Nash equilibrium framework to optimize the seed selection strategy. This framework aims to balance the opponent's potential payoffs as evenly as possible while preserving the stochasticity of the focal player's strategy, thereby providing theoretical guidance for the agent to make optimal rational decisions under conditions of uncertainty.

4.3. Parameter analysis

4.3.1. Impact of σ

Experiments were performed on four real-world datasets, with the propagation duration fixed at 50 and the number of seeds varied from 0 to 50. The influence probability σ was configured as either a fixed value (i.e., 0.2) or a random value, to investigate the variations in the propagation performance of the DM-DDQN algorithm. The results are presented in Figure 8. It is worth noting that a TRIVALENT model [56] was employed to generate random values in this work, where the influence probability σ for each node is uniformly sampled from the set $\{0.2, 0.05, 0.01\}$, corresponding to high, medium, and low levels, respectively.

As illustrated in Figure 8, when the individual influence probability σ varies, the propagation performance of our proposed DM-DDQN exhibits slight fluctuations, yet its impact on the overall opinion evolution is negligible. Specifically, the number of seed nodes plays a more critical role in guiding individual opinions within the population compared to the stubbornness probability. Thus, in practical rumor propagation control, even in networks with high stubbornness probabilities, optimizing the truth source selection strategy can effectively improve the intervention effect on network opinions and, to a certain extent, steer the future development direction of public opinion.

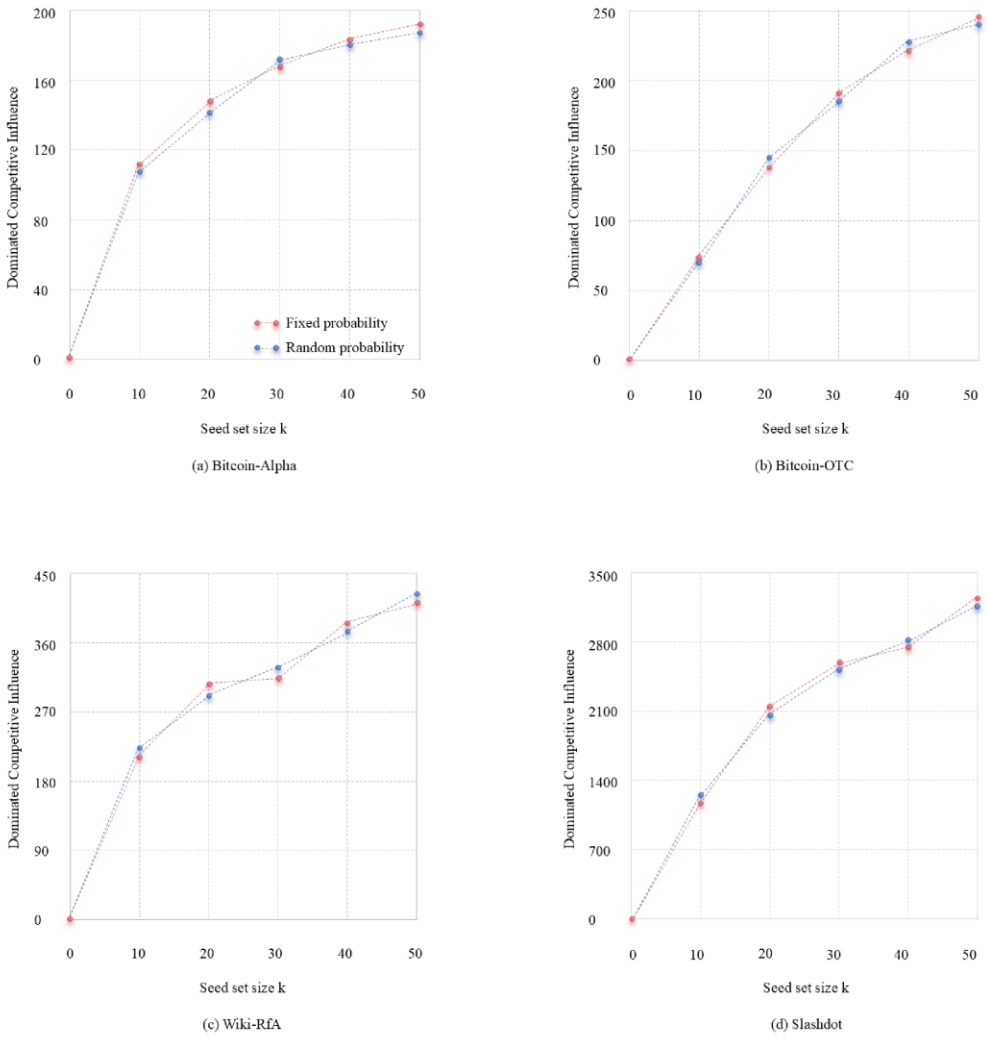

Figure 8. Performance comparison under different values of σ.

4.3.2. Impact of φ

Experiments were performed on four real-world datasets, where the number of seeds and propagation duration were fixed at 50. The optimal value of the initial learning rate φ was searched within the range [0.0001, 0.002] to maximize the performance of the proposed DM-DDQN algorithm. The results are presented in Figure 9. As illustrated in Figure 9, the optimal value of the initial learning rate φ is 0.001, at which the propagation performance of the proposed DM-DDQN attains its maximum. This finding indicates that the adoption of the PER and soft update mechanisms for network parameter updating effectively mitigates the sensitivity to the initial learning rate during the training process. Thus, a moderate relaxation of the constraints on the initial learning rate can accelerate the convergence rate without inducing training oscillations. Notably, even though the Adam optimizer is employed in this paper, the initial learning rate still exerts a non-negligible influence on the training outcomes.

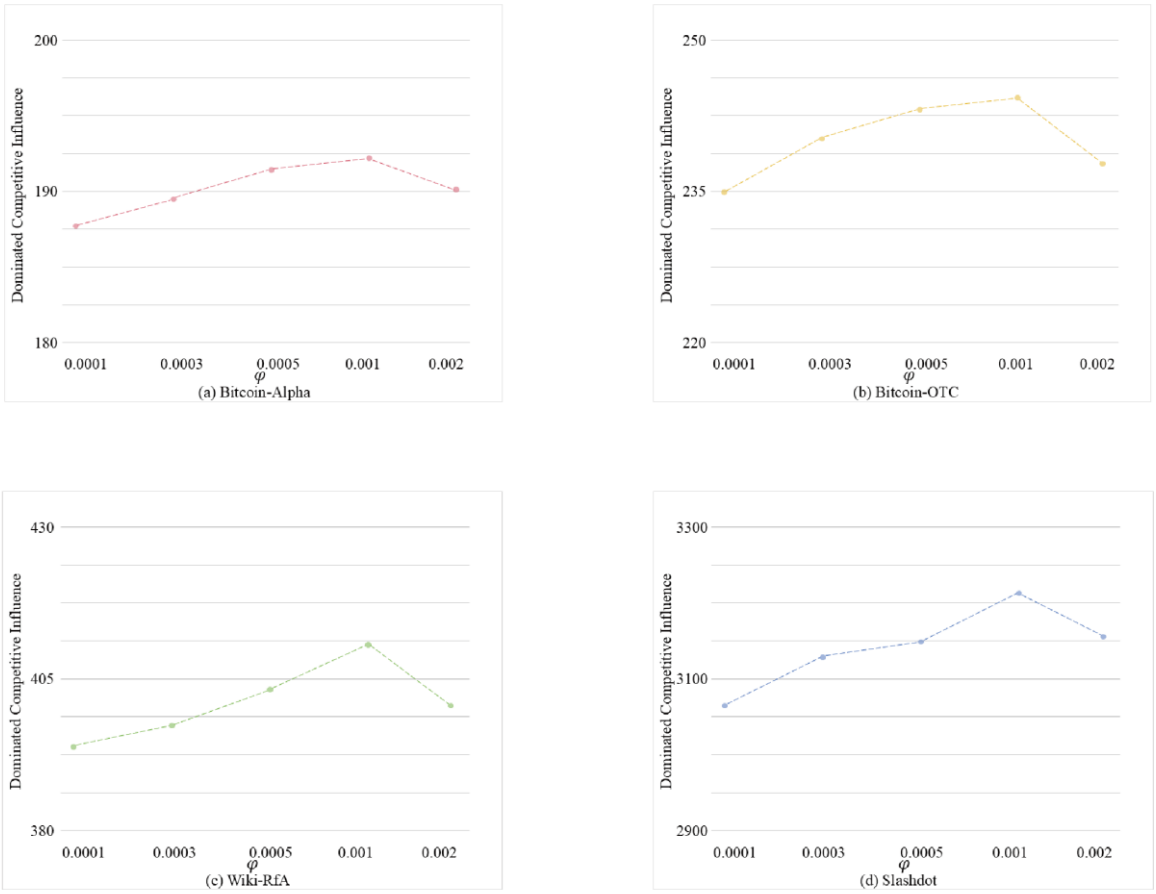

Figure 9. Performance comparison under different values of φ.

4.4. Training episodes and time performance

This paper fixes the number of seeds at 10 and the time step at 50. The optimal number of training episodes, training time, and running time of the DM-DDQN algorithm are shown in Table 5.

Table 5. Best training episodes and time performance.

Networks	Training episodes	Training time	Running time
Bitcoin-Alpha	2100	43 min	9 s
Bitcoin-OTC	2600	71 min	15 s
Wiki-RfA	3400	2.1 h	32 s
Slashdot	5200	5.7 h	108 s

The overall complexity of the proposed DM-DDQN algorithm is relatively high. In future work, we aim to significantly reduce computational costs to make it scalable for large-scale social networks and further investigate delayed competition diffusion in dynamic networks to broaden its application scenarios.

5. Conclusions

This paper addresses the challenge of online rumor governance. First, we formulate the time-sensitive dominant competition influence maximization problem (i.e., TDCIM), which aims to determine the optimal placement of truth sources such that truth can capture the high ground of public opinion and effectively curb rumor propagation. Second, by integrating the echo chamber effect and structural balance theory, we optimize the rules for opinion updating and competition and propose a continuous opinion competition diffusion model (i.e., COC) tailored for directed signed social networks. Third, leveraging directed signed graph embedding and mixed-strategy Nash equilibrium, we reconstruct the state and action spaces of DDQN and develop a truth source selection algorithm based on game theory (i.e., DM-DDQN). Finally, experiments are conducted on four real-world online social networks to analyze the impact of the number and duration of truth source deployments on rumor suppression. These findings enable decision-makers to flexibly adjust resource allocation strategies under different budget constraints, thereby achieving efficient utilization of governance resources and maximizing governance effectiveness.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (2024JKF01).

Conflict of interest

The authors declare there is no conflict of interest.

Author contributions

Ying Guo: Conceptualization, data curation, formal analysis, investigation, methodology, software, and writing-original draft. Yong-nan Li: Funding acquisition, project administration, resources, supervision, validation, visualization, and writing-review and editing.

References

1. Y. Guo, Y. N. Li, Multiple rumor source localization in signed social networks driven by data intelligence, *AIMS Math.*, **11** (2026), 1927–1953. https://doi.org/10.3934/math.2026080
2. W. Zhu, Y. Huang, S. Yang, Y. Miao, C. Peng, A positive influence maximization algorithm in signed social networks, *Comput. Mater. Contin.*, **76** (2023), 1977–1994. https://doi.org/10.32604/cmc.2023.040998

3. C. Dong, G. Q. Xu, L. Meng, CRB: A new rumor blocking algorithm in online social networks based on competitive spreading model and influence maximization, *Chin. Phys. B*, **33** (2024), 088901. https://doi.org/10.1088/1674-1056/ad531f
4. D. Li, Y. Wang, M. Li, X. Sun, J. Pan, J. Ma, Net positive influence maximization in signed social networks, *J. Intell. Fuzzy Syst.*, **41** (2021), 3821–3832. https://doi.org/10.3233/JIFS-191908
5. S. Kumar, A. Mallik, B. S. Panda, Influence maximization in social networks using transfer learning via graph-based LSTM, *Expert Syst. Appl.*, **212** (2023), 118770. https://doi.org/10.1016/j.eswa.2022.118770
6. L. Ma, Z. Shao, X. Li, Q. Lin, J. Li, V. C. M. Leung, et al., Influence maximization in complex networks by using evolutionary deep reinforcement learning, *IEEE Trans. Emerg. Top. Comput. Intell.*, **7** (2023), 995–1009. https://doi.org/10.1109/TETCI.2021.3136643
7. C. Guo, W. Li, F. Liu, K. Zhong, X. Wu, Y. Zhao, et al., Influence maximization algorithm based on group trust and local topology structure, *Neurocomputing*, **564** (2024), 126936. https://doi.org/10.1016/j.neucom.2023.126936
8. W. Zhu, K. Zhang, J. Zhong, C. Hou, J. Ji, BiGDN: An end-to-end influence maximization framework based on deep reinforcement learning and graph neural networks, *Expert Syst. Appl.*, **270** (2025), 126384. https://doi.org/10.1016/j.eswa.2025.126384
9. H. Li, L. Pan, P. Wu, Dominated competitive influence maximization with time-critical and time-delayed diffusion in social networks, *J. Comput. Sci.*, **28** (2018), 318–327. https://doi.org/10.1016/j.jocs.2017.10.015
10. A. I. E. Hosni, K. Li, C. Ding, S. Ahmed, Least cost rumor influence minimization in multiplex social networks, in *Proceedings of the 25th International Conference on Neural Information Processing*, (2018), 93–105. https://doi.org/10.1007/978-3-030-04224-0_9
11. D. T. Nguyen, H. Zhang, S. Das, M. T. Thai, T. N. Dinh, Least cost influence in multiplex social networks: Model representation and analysis, in *Proceedings of the 13th International Conference on Data Mining*, (2013), 567–576. https://doi.org/10.1109/ICDM.2013.24
12. Q. He, X. Wang, Y. Zhao, B. Yi, X. Lu, M. Yang, M. Huang, Reinforcement-learning-based competitive opinion maximization approach in signed social networks, *IEEE Trans. Comput. Soc. Syst.*, **9** (2022), 1505–1514. https://doi.org/10.1109/TCSS.2021.3120421
13. L. Yu, T. Yang, C. Yao, D. Fu, Y. Wang, Opinion maximization on social trust networks based on game theory and DQN method, *Inf. Sci.*, **729** (2026), 122879. https://doi.org/10.1016/j.ins.2025.122879
14. Q. He, L. Zhang, H. Fang, X. Wang, L. Ma, K. Yu, et al., Multistage competitive opinion maximization with Q-learning-based method in social networks, *IEEE Trans. Neural Netw. Learn. Syst.*, **36** (2025), 7158–7168. https://doi.org/10.1109/TNNLS.2024.3387293
15. A. Hevapathige, Q. Wang, A. N. Zehmakan, DeepSN: A sheaf neural framework for influence maximization, in *Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence*, (2025), 17177–17185. https://doi.org/10.1609/aaai.v39i16.33888
16. W. Zhu, K. Zhang, Y. Bai, A. Yu, Deep reinforcement learning-based algorithm for maximizing the influence of location-based social networks, *Neurocomputing*, **671** (2026), 132747. https://doi.org/10.1016/j.neucom.2026.132747
17. Y. Zou, G. Li, Y. Wang, B. Ning, S. Wang, Topic-aware influence maximization based on deep reinforcement learning and group relative optimization, *Appl. Soft Comput.*, **192** (2026), 114786. https://doi.org/10.1016/j.asoc.2026.114786

18. Y. Wang, Y. Zhang, F. Yang, D. Li, X. Sun, J. Ma, Time-sensitive positive influence maximization in signed social networks, *Phys. A*, **584** (2021), 126353. https://doi.org/10.1016/j.physa.2021.126353
19. Q. He, H. Du, Z. Liang, Positive influence maximization in signed networks within a limited time, *IEEE Trans. Comput. Soc. Syst.*, **10** (2023), 2624–2635. https://doi.org/10.1109/TCSS.2022.3192410
20. C. E. Sung, H. S. Ma, J. W. Huang, Positive influence maximization and negative influence minimization in signed networks under competitive independent cascade model, in *Proceedings of the 7th International Conference on Data Science and Advanced Analytics*, (2020), 236–244. https://doi.org/10.1109/DSAA49011.2020.00036
21. Y. Li, H. Gao, Y. Gao, J. Guo, W. Wu, A survey on influence maximization: From an ML-based combinatorial optimization, *ACM Trans. Knowl. Discov. Data*, **17** (2023), 133. https://doi.org/10.1145/3604559
22. H. Li, M. Xu, S. S. Bhowmick, J. S. Rayhan, C. Sun, J. Cui, PIANO: Influence maximization meets deep reinforcement learning, *IEEE Trans. Comput. Soc. Syst.*, **10** (2023), 1288–1300. https://doi.org/10.1109/TCSS.2022.3164667
23. T. Chen, S. Yan, J. Guo, W. Wu, ToupleGDD: A fine-designed solution of influence maximization by deep reinforcement learning, *IEEE Trans. Comput. Soc. Syst.*, **11** (2024), 2210–2221. https://doi.org/10.1109/TCSS.2023.3272331
24. J. Wang, Z. Cao, C. Xie, Y. Li, J. Liu, Z. S. Gao, DGN: Influence maximization based on deep reinforcement learning, *J. Supercomput.*, **81** (2025), 130. https://doi.org/10.1007/s11227-024-06621-9
25. D. Kempe, J. Kleinberg, É. 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*, (2003), 137–146. https://doi.org/10.1145/956750.956769
26. E. Yildiz, A. Ozdaglar, D. Acemoglu, A. Saberi, A. Scaglione, Binary opinion dynamics with stubborn agents, *ACM Trans. Econ. Comput.*, **1** (2013), 19. https://doi.org/10.1145/2538508
27. L. Vassio, F. Fagnani, P. Frasca, A. Ozdaglar, Message passing optimization of harmonic influence centrality, *IEEE Trans. Control Netw. Syst.*, **1** (2014), 109–120. https://doi.org/10.1109/TCNS.2014.2304870
28. M. Bini, P. Frasca, C. Ravazzi, F. Dabbene, Graph structure-based heuristics for optimal targeting in social networks, *IEEE Trans. Control Netw. Syst.*, **9** (2022), 1189–1201. https://doi.org/10.1109/TCNS.2022.3163665
29. Q. Zhang, D. Chen, L. M. Kaplan, A. Jøsang, D. H. Jeong, F. Chen, et al., Beyond binary opinions: A deep reinforcement learning-based approach to uncertainty-aware competitive influence maximization, *IEEE Trans. Netw. Sci. Eng.*, **13** (2026), 2219–2235. https://doi.org/10.1109/TNSE.2025.3611330
30. F. K. Chang, S. C. Li, J. W. Huang, Efficient influence maximization in signed networks with positive influence increasing and negative influence decreasing simultaneously via edge-view influence estimation, in *Social Networks Analysis and Mining, ASONAM 2025, Lecture Notes in Computer Science, Springer, Cham*, **16322** (2026), 142–150. https://doi.org/10.1007/978-3-032-13513-1_12

31. Y. Liu, Y. Zhong, X. Li, P. Zhu, Z. Wang, Vital nodes identification via evolutionary algorithm with percolation optimization in complex networks, *IEEE Trans. Netw. Sci. Eng.*, **11** (2024), 3838–3850. https://doi.org/10.1109/TNSE.2024.3388994
32. N. Wang, W. Xie, V. Tiberius, Y. Qiu, Accelerating new product diffusion: How lead users serve as opinion leaders in social networks, *J. Retail. Consum. Serv.*, **72** (2023), 103297. https://doi.org/10.1016/j.jretconser.2023.103297
33. S. Bharathi, D. Kempe, M. Salek, Competitive influence maximization in social networks, in *Internet and Network Economics, WINE 2007, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg,* **4858** (2007), 306–311. https://doi.org/10.1007/978-3-540-77105-0_31
34. X. He, G. Song, W. Chen, Q. Jiang, Influence blocking maximization in social networks under the competitive linear threshold model, in *Proceedings of the 2012 SIAM International Conference on Data Mining*, (2012), 463–474. https://doi.org/10.1137/1.9781611972825.40
35. N. E. Friedkin, E. C. Johnsen, Social influence and opinions, *J. Math. Sociol.*, **15** (1990), 193–206. https://doi.org/10.1080/0022250X.1990.9990069
36. H. Zhang, Y. Li, Y. Chen, H. V. Zhao, Smart evolution for information diffusion over social networks, *IEEE Trans. Inf. Forensics Secur.*, **16** (2021), 1203–1217. https://doi.org/10.1109/tifs.2020.3032039
37. F. Heider, Attitudes and cognitive organization, *J. Psychol.*, **21** (1946), 107–112. https://doi.org/10.1080/00223980.1946.9917275
38. D. Cartwright, F. Harary, Structural balance: A generalization of Heider's theory, *Psychol. Rev.*, **63** (1956), 277–293. https://doi.org/10.1037/h0046049
39. D. Choi, S. Chun, H. Oh, J. Han, T. Kwon, Rumor propagation is amplified by echo chambers in social media, *Sci. Rep.*, **10** (2020), 310. https://doi.org/10.1038/s41598-019-57272-3
40. M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media, *Proc. Natl. Acad. Sci. U.S.A.*, **118** (2021), e2023301118. https://doi.org/10.1073/pnas.2023301118
41. P. Cisneros-Velarde, K. S. Chan, F. Bullo, Polarization and fluctuations in signed social networks, *IEEE Trans. Automat. Contr.*, **66** (2021), 3789–3793. https://doi.org/10.1109/TAC.2020.3024967
42. G. Shi, C. Altafini, J. S. Baras, Dynamics over signed networks, *SIAM Rev.*, **61** (2019), 229–257. https://doi.org/10.1137/17M1134172
43. Q. He, L. Sun, X. Wang, Z. Wang, M. Huang, B. Yi, et al., Positive opinion maximization in signed social networks, *Inf. Sci.*, **558** (2021), 34–49. https://doi.org/10.1016/j.ins.2020.12.091
44. Y. Liu, Y. Hu, S. Yu, X. Zhou, K. Li, Conformity-aware adoption maximization in competitive social networks, *Neurocomputing*, **573** (2024), 127224. https://doi.org/10.1016/j.neucom.2023.127224
45. H. van Hasselt, A. Guez, D. Silver, Deep reinforcement learning with double Q-learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **30** (2016), 2094–2100. https://doi.org/10.1609/aaai.v30i1.10295
46. S. C. Lin, S. D. Lin, M. S. Chen, A learning-based framework to handle multi-round multi-party influence maximization on social networks, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2015), 695–704. https://doi.org/10.1145/2783258.2783392

47. Y. Xiao, G. Kang, J. Liu, B. Cao, L. Ding, WSGCN4SLP: Weighted signed graph convolutional network for service link prediction, in *Proceedings of the 2021 IEEE International Conference on Web Services*, (2021), 135–144. https://doi.org/10.1109/ICWS53863.2021.00029
48. Y. Wu, L. Hu, Y. Wang, Signed attention based graph neural network for graphs with heterophily, *Neurocomputing*, **557** (2023), 126731. https://doi.org/10.1016/j.neucom.2023.126731
49. F. Morone, H. A. Makse, Influence maximization in complex networks through optimal percolation, *Nature*, **524** (2015), 65–68. https://doi.org/10.1038/nature14604
50. Z. Zhao, X. Wang, W. Zhang, Z. Zhu, A community-based approach to identifying influential spreaders, *Entropy*, **17** (2015), 2228–2252. https://doi.org/10.3390/e17042228
51. J. Nash, Non-cooperative games, *Ann. Math.*, **54** (1951), 286–295. https://doi.org/10.2307/1969529
52. T. Schaul, J. Quan, I. Antonoglou, D. Silver, Prioritized experience replay, in *Proceedings of the 2016 International Conference on Learning Representations*, (2016). Available from: https://mlanthology.org/iclr/2016/schaul2016iclr-prioritized/.
53. S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, V. S. Subrahmanian, REV2: Fraudulent user prediction in rating platforms, in *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, (2018), 333–341. https://doi.org/10.1145/3159652.3159729
54. R. West, H. S. Paskov, J. Leskovec, C. Potts, Exploiting social network structure for person-to-person sentiment analysis, *Trans. Assoc. Comput. Linguist.*, **2** (2014), 297–310. https://doi.org/10.1162/tacl_a_00184
55. J. Leskovec, D. Huttenlocher, J. Kleinberg, Signed networks in social media, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2010), 1361–1370. https://doi.org/10.1145/1753326.1753532
56. W. Chen, C. Wang, Y. Wang, 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*, (2010), 1029–1038. https://doi.org/10.1145/1835804.1835934



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)