applied sciences

MDPI

Article

Adaptive Boundary-Aware Fact-Checker Placement for Misinformation Suppression in Social Networks

Mostafa Taghizade Firouzjaee [1], Ghazal Naderi [1], Ross Gore [2] and Neda Moghim [2,3,*]

[1] Department of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol 47148-71167, Iran; taghizade.mostafa@nit.ac.ir (M.T.F.); ghazalnaderi4@gmail.com (G.N.)
[2] Center for Secure and Intelligent Critical Systems, Old Dominion University, Suffolk, VA 23435, USA; rgore@odu.edu
[3] School of Cybersecurity, Old Dominion University, Norfolk, VA 23529, USA
* Correspondence: nmoghim@odu.edu

Abstract

The spread of fake news on online social networks is driven by imitation-based user behavior and network topology, often leading to persistent misinformation clusters and echo chambers. In this study, we develop a spatial evolutionary game-theoretic framework in which agents update their latent opinions through payoff-biased imitation, while external fact-checkers act as non-imitative intervention nodes. Building on this formulation, we propose an adaptive, boundary-aware intervention mechanism that dynamically regulates both the density and spatial allocation of fact-checkers according to real-time system conditions. Competing information clusters are identified through local neighborhood composition, enabling boundary nodes, i.e., interfaces between fake-news and non-fake-news regions, to be detected and targeted where strategic shifts are most likely to occur. Importantly, fact-checking is modeled as an external intervention that may induce a probabilistic lasting correction on agents' latent opinions after removal, capturing more realistic post-intervention behavior. Unlike static strategies that assume fixed fact-checker distributions, the proposed approach continuously reallocates interventions toward structurally critical regions, while adaptively adjusting resource intensity based on misinformation prevalence. Extensive simulations on small-world, scale-free, and random networks show that the adaptive model consistently outperforms static baselines, reducing the final fake-news prevalence by over 90%, accelerating suppression, and improving overall system efficiency. Statistical tests confirm the significance of these improvements ($p < 0.001$), while sensitivity analyses demonstrate robustness across parameter settings and intervention assumptions.

Keywords: fake news suppression; spatial evolutionary game theory; adaptive intervention; boundary-aware placement; social networks

Check for updates

Academic Editor: Jose María Alvarez Rodríguez

Received: 8 April 2026
Revised: 1 May 2026
Accepted: 4 May 2026
Published: 11 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Recent real-world crises, from the COVID-19 infodemic to coordinated disinformation during elections, underscore the urgency for effective misinformation control. Social media platforms enable rapid dissemination of content across vast audiences. However, this capability fuels the explosive spread of misinformation. Empirical studies reveal that false news travels significantly farther, faster, and deeper than truthful information [1]. This dynamic fosters feedback loops of selective exposure, where users with similar beliefs cluster together, reinforcing shared narratives and forming so-called "echo chambers" [2]. In

an attention-driven digital ecosystem, these echo chambers not only sustain misinformation but also weaken the effectiveness of traditional fact-checking interventions.

Evolutionary game theory provides a natural framework for modeling such dynamics, where agents' latent opinions evolve over time through payoff-biased imitation, while external interventions may temporarily influence observable behavior. This framework allows researchers to simulate how different types of news propagate and interact in structured networks. The growing impact of fake news, coupled with the structural resilience provided by network clustering, motivates the need for intervention strategies that are adaptive to the evolving state of the system and sensitive to its topology.

Prior research has modeled fake news dissemination using spatial evolutionary games with a fixed complement of fact-checkers. For example, Jones et al. [3] introduced a game-theoretic model in which crowdsourced fact-checkers impose penalties on fake news spreaders and rewards on real news agents. They showed that even when fact-checkers are present, local payoff-biased imitation combined with clustered network neighborhoods causes fake news agents to reinforce each other's strategies, leading to the self-organization of tightly knit echo chambers that enhance resilience to misinformation. They also demonstrated that carefully choosing which individuals to equip as fact-checkers can greatly improve performance. However, these models typically assume that the number and placement of fact-checkers remain fixed over time without adapting to the evolving prevalence of fake news. In reality, misinformation outbreaks may occur suddenly or surge in particular communities, suggesting the need for dynamic, network-aware intervention policies.

In this work, we assume that the type of agent (truthful or fake-news spreader) can be identified by an external monitoring system. This assumption reflects the availability of imperfect but operational detection mechanisms (e.g., platform moderation tools or automated classifiers), and allows us to focus on the suppression dynamics rather than detection accuracy. We introduce a novel real-time adaptive intervention strategy in which the density and spatial allocation of fact-checkers evolve in response to the current state of the system. Fact-checkers are strategically placed at the boundaries between competing information clusters, i.e., interfaces separating regions dominated by truthful (A) and fake-news (B) agents, where agents are most susceptible to switching strategies. Moreover, fact-checking is modeled as an external intervention that may induce a probabilistic lasting correction on agents' latent opinions after the intervention is removed, capturing more realistic post-exposure behavioral effects.

Our key contributions are summarized as follows:

- We develop a spatial evolutionary game model in which latent opinions evolve through payoff-biased imitation, while fact-checkers act as an external intervention layer that dynamically interacts with the underlying opinion dynamics.
- We design a boundary-aware placement mechanism that identifies structurally critical interface nodes between competing information clusters and dynamically allocates fact-checkers to these regions where behavioral transitions are most likely.
- We incorporate a probabilistic lasting correction mechanism, allowing fact-checking interventions to have persistent but non-deterministic effects on agents' latent opinions, thereby improving behavioral realism.
- We evaluate our method through comprehensive simulations on multiple network topologies, demonstrating that adaptive allocation of fact-checkers significantly improves suppression performance compared to static baselines.

Our results indicate that the proposed adaptive structure-based intervention drives down the fake-news population more rapidly and to significantly lower steady-state levels compared to non-adaptive baselines, while maintaining stability and robustness across different network structures. These findings emphasize that accounting for network

structure and real-time system state can substantially enhance misinformation control. Given that the viral spread of misinformation is now regarded as one of the most serious threats to public discourse [4], our adaptive boundary-aware approach offers a promising step toward more effective targeted mitigation strategies in social media environments.

The remainder of this paper is organized as follows. Section 2 reviews related work on misinformation diffusion and intervention strategies in networked systems. Section 3 presents the proposed spatial evolutionary game framework and the adaptive boundary-aware intervention mechanism. Section 4 describes the experimental setup and simulation results across different network topologies. Section 5 provides a discussion of the results, and Section 6 concludes the paper with future research directions.

2. Related Work

The rapid spread of misinformation on social networks has been extensively documented. Empirical analyses reveal that false information often propagates faster and deeper than truthful content, a phenomenon amplified by the emergence of echo chambers and tightly knit communities where like-minded users preferentially share similar narratives [1,5]. Researchers have applied network-structured dynamic models to model these effects. Spatial and evolutionary game-theoretical (EGT) frameworks have been particularly fruitful, as they explicitly model local imitation dynamics and strategic interactions on network structures, which are central to misinformation propagation. For example, Hsu et al. [6] formulated a game-theoretic model in which users imitate successful neighbors, showing that network topology can enable self-reinforcing clusters of misinformation. Meng et al. [7] likewise treated misinformation as a competing species in a cooperative game, finding that high misinformation prevalence undermines collective cooperation. Such spatial EGT and epidemic-inspired models emphasize how local strategic imitation and network clusters can entrench false narratives, thereby highlighting the structural drivers of misinformation.

Another strand of work focuses on intervention strategies within these structural models. Classic approaches often assume a fixed set of fact-checker or immune nodes that impose penalties on misinformation spreaders. Liu et al. [8] developed a tri-level game-theoretic model for governments, platforms, and users, showing that raising penalties for sharing false content or lowering its payoff can significantly reduce misinformation. Jones et al. [3] similarly showed that simply adding static fact-checkers is often insufficient; network echo chambers still enable misinformation to persist unless interventions are topologically aware. These static schemes, whether based on node centrality or fixed policies, do not adapt as the cascade evolves and often assume that interventions directly overwrite agent states, rather than acting as a temporary external layer influencing latent opinions, neglecting the distinction between latent opinions and external interventions. Generally, this line of work relates to the classic influence-minimization problem in networks. However, selecting the best nodes to immunize or block is computationally hard (NP-hard) [9,10]; therefore, many methods rely on heuristics or greedy approximations. These static schemes fail when misinformation surges dynamically.

Recent studies have explored adaptive and control-based interventions. For instance, Govindankutty and Gopalan [11] used an epidemic model to study state-dependent mitigation: interventions (blockers, moderators) are placed dynamically in response to real-time rumor prevalence, which improves the containment of outbreaks. Acemoglu et al. [12] studied filtering algorithms and showed that personalization (filtering quality) can ironically increase ideological segregation and misinformation cascades under low information quality. Other theoretical models derive optimal control strategies. For example, Jing et al. [13] formulated a rumor propagation model and derived optimal control strategies to

suppress rumor spreading at minimum cost. Similarly, reinforcement-learning methods have been proposed to adaptively block rumors: Jiang et al. [14] define a Dynamic Rumor Influence Minimization (DRIM) problem and use deep Q-learning to select blockers at each step, outperforming static centrality-based heuristics. These dynamic approaches can tailor interventions to the evolving network state; however, they typically require detailed knowledge or costly training (e.g., reinforcement learning policies depend on global state information and repeated retraining), and they do not explicitly model how interventions affect agents' underlying opinions after removal, nor do they account for potentially persistent but non-deterministic post-intervention effects.

However, most of the modern data-driven literature on misinformation uses a classification approach. Machine learning and deep models, often operating on content, social contexts, or network features, aim to distinguish fake from true news posts. For example, a recent review described dozens of fake news classifiers based on linguistic, visual, and propagation features [15]. Graph neural networks (GNNs) have been applied to fuse textual and network information, and Phan et al. [16] provided a comprehensive survey of GNN-based fake-news detection, highlighting how GNNs capture complex relational features. In practice, these methods can achieve high detection accuracy given ample training data, but they have notable limitations. They rely on supervised learning with large labeled datasets, which are often scarce or biased in emerging misinformation events. They can overfit or underperform when the data are limited or rapidly changing. Moreover, classification models are intrinsically reactive; they flag individual posts or users but do not directly prescribe how to intervene in the network to halt propagation. Advanced approaches such as deep RL can adaptively choose interventions but at the cost of heavy computation and reliance on accurate models of user behavior. In addition, these methods rarely capture the dynamic feedback between intervention actions and evolving user behavior, which is critical for effective suppression.

In summary, existing studies on misinformation fall into two main categories. Structure-based models (e.g., EGT, epidemic, and control-theoretic approaches) emphasize network topology and strategic interactions but often rely on static or globally optimized interventions that are difficult to implement in practice. Data-driven approaches (e.g., machine learning and graph neural networks) provide strong detection capabilities but are limited by data requirements, potential overfitting, and the lack of explicit control mechanisms for intervention. Our study bridges these domains by proposing a spatially informed adaptive intervention framework. Specifically, we extend EGT-based models by introducing (i) a feedback-driven adaptation of fact-checker density based on misinformation prevalence, (ii) a boundary-aware placement mechanism that targets structurally critical interface nodes, and (iii) a probabilistic lasting correction mechanism that captures persistent but non-deterministic effects of fact-checking on agents' latent opinions. This combination allows our approach to remain both computationally efficient and behaviorally realistic, while capturing the interaction between network structure, adaptive control, and user-level behavioral dynamics, thereby overcoming the limitations of static and purely data-driven methods.

3. Methodology

In this section, we first define the system model and the agent types underlying the spatial evolutionary game framework. We then describe the evolutionary interaction dynamics and introduce the proposed adaptive intervention mechanism, which combines feedback-based regulation of fact-checker density with boundary-aware spatial placement to suppress the spread of misinformation. The proposed framework models information diffusion in a networked population composed of latent opinion states and an external

intervention layer. Specifically, agents hold latent opinions corresponding to real news spreaders (A) or fake news spreaders (B), while fact-checkers (C) are modeled as externally assigned intervention nodes that temporarily override observable behavior without directly replacing latent opinions. We assume that the type of agent (truthful or fake-news spreader) is observable to the monitoring mechanism. In other words, the system can identify fake-news spreaders without classification errors, allowing us to focus on intervention and suppression dynamics rather than detection uncertainty. Agents interact locally and update their latent opinions through payoff-driven evolutionary dynamics, whereas fact-checkers act as externally assigned intervention nodes that influence interactions during their deployment. The following subsections formalize the graph representation, interaction rules, and adaptive control mechanism in detail.

3.1. Graph-Theoretic Representation and Agent Types

The interaction structure of the system is modeled as a graph $G = (V, E)$, where nodes $v_i \in V$ represent individual agents, and edges $e_{ij} \in E$ denote bilateral social interactions through which information is exchanged. Each agent located at node i adopts latent opinion $o_i(t) \in \{A, B\}$ at discrete time t, while fact-checking is represented as an external intervention layer rather than an intrinsic strategy. Strategies A and B correspond to opinion-bearing agents who disseminate truthful information and misinformation, respectively. In contrast, fact-checkers (C) are not treated as a third endogenous strategy, but as externally assigned intervention nodes that temporarily override observable behavior without directly modifying latent opinions. Under evolutionary imitation dynamics, agents periodically compare their payoffs with those of neighboring agents and may adopt latent opinions associated with higher fitness. Fact-checkers do not participate in opinion propagation or imitation-based updates; instead, they act as exogenous moderators that locally influence interactions without entering the evolutionary state space. As the fact-checker state is externally assigned rather than endogenously learned, it does not evolve through imitation, and nodes revert to their latent opinions once the intervention is removed (subject to probabilistic lasting correction effects introduced later). Nevertheless, the set of nodes designated as fact-checkers may change over time through adaptive reallocation. This modeling choice captures two essential features of real online platforms. First, user behavior evolves through localized social learning, in which individuals update their latent opinions by imitating more successful or influential neighbors. Second, moderation mechanisms, such as fact-checks, warnings, or platform-level interventions, are imposed exogenously and do not emerge from user imitation. Thus, modeling fact-checkers as an external intervention layer rather than a standard strategy allows a clearer separation between endogenous opinion dynamics and exogenous control mechanisms. Overall, this framework provides a flexible yet interpretable representation of decentralized information dynamics, consistent with established models of cultural evolution and behavioral learning in networks [6,7], while explicitly incorporating targeted external interventions.

3.2. Evolutionary Game Dynamics and Local Interactions

Agents repeatedly engage in pairwise interactions with immediate neighbors. Each interaction yields a payoff governed by the following payoff matrix:

$$\mathrm{P} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -4 \\ 0 & 0 & 0 \end{pmatrix},$$

where strategies correspond to A (truthful agents), B (fake news spreaders), and C (fact checkers), respectively. The payoff of each agent is determined by its own strategy and that

of its interaction partner. This payoff structure is designed to capture the key incentives underlying information propagation dynamics. The interactions between truthful agents (A–A) yield a modest mutual benefit (payoff 1), reflecting the reinforcement of accurate information. By contrast, fake news spreaders (B) receive a higher payoff (2) when interacting with others of the same type, representing the amplification effects observed in misinformation echo chambers [4,5]. However, when exposed to fact-checkers, fake news agents incur a substantial penalty (-4), modeling the corrective impact of fact checking combined with reduced visibility or credibility loss. Interactions between truthful agents and fact-checkers (A–C) provide a weakly positive reinforcement, while fact-checkers themselves do not propagate opinions and therefore receive zero payoffs in all interactions. The numerical payoff values used here are chosen to reflect the intended incentive structure of the model. Alternative payoff settings that preserve the same relative strategic relationships (e.g., stronger reinforcement among fake news agents and penalties under fact checking) are expected to produce qualitatively similar evolutionary dynamics.

The cumulative payoff of agent i at time t is given by:

$$\pi_i(t) = \sum_{j \in \mathcal{N}(i)} P\big(s_i(t), s_j(t)\big) \tag{1}$$

where $\mathcal{N}(i)$ denotes the set of neighbors of node i, and $s_i(t)$ represents the effective strategy of agent i, determined by its latent opinion and its possible assignment as a fact-checker.

To model evolutionary pressure, payoffs are mapped to fitness using an exponential transformation:

$$f_i(t) = \exp(\beta\, \pi_i(t)), \tag{2}$$

where $\beta \geq 0$ controls the selection intensity. In the limit $\beta \to 0$, strategy updates approach neutral drift, whereas larger values of β induce increasingly deterministic imitation of higher payoff strategies. At each asynchronous update step, a non-fact-checker agent, i is selected uniformly at random. Agent i then selects one neighbor $j \in \mathcal{N}(i)$ with a probability proportional to $f_j(t)$ and adopts the neighbor's latent opinion. Fact-checker nodes are excluded from the imitation process and do not update their state through evolutionary dynamics. This update rule follows pairwise comparison dynamics [6], adapted here to distinguish between latent opinion evolution and externally imposed intervention states.

Figure 1 illustrates the adaptive boundary-aware intervention framework. The model operates through two interacting loops: (i) a lower loop governing the core evolutionary dynamics, in which agents accumulate local payoffs, convert them into fitness values, and update their latent opinions by imitating higher-payoff neighboring agents (payoff-biased imitation), excluding fact-checkers; and (ii) an upper control loop, activated at every (T) time step that monitors the current prevalence of fake news ($\rho(t)$), adaptively adjusts the fact-checker ratio, detects misinformation cluster boundaries, and reallocates fact-checkers to strategically important interface nodes identified based on local neighborhood composition. In this way, the system integrates real-time feedback with network structural awareness, enabling efficient and robust suppression of misinformation under evolving conditions.

3.3. Adaptive Fact-Checker Regulation

To capture the inherently dynamic nature of misinformation diffusion, we introduce a feedback-driven regulation mechanism that adaptively adjusts the density of fact checkers in response to the observed prevalence of fake news. Let $n_A(t)$, $n_B(t)$, and $n_C(t)$ denote the number of truthful agents (A), fake news spreaders (B), and fact-checkers (C) at time t, respectively, where fact-checkers represent externally assigned intervention nodes rather than intrinsic agent types.

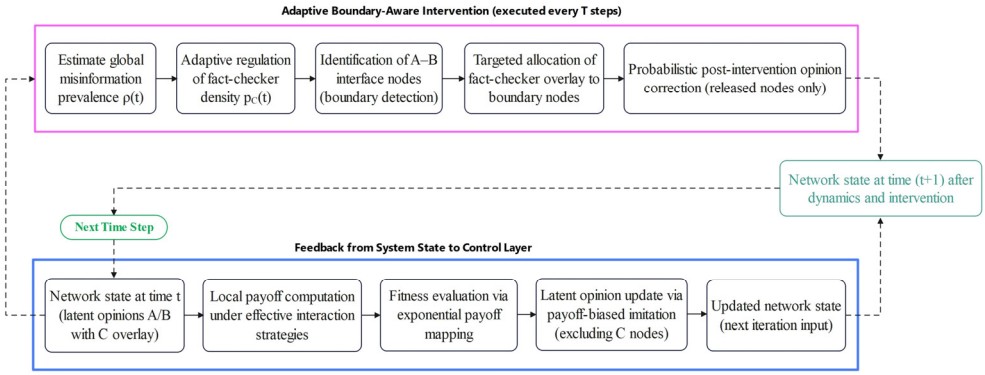


Figure 1. Flowchart of the adaptive boundary-aware intervention framework. The upper pink panel depicts the adaptive control loop (every T steps) and the lower blue panel shows the evolutionary dynamics. Solid arrows represent main flow; dashed arrows indicate feedback and transitions.

We define the instantaneous prevalence of misinformation as:

$$p_B(t) = \frac{n_B(t)}{n_A(t) + n_B(t)} \tag{3}$$

which measures the relative dominance of fake-news agents among latent opinion-bearing nodes. This normalization excludes fact-checkers by design, since they act as an external intervention layer and do not participate in opinion propagation.

The fact-checker density $p_C(t)$ is then determined using a clipped affine feedback rule:

$$p_C(t) = \text{clip}\Big(p_{\min} + k_{\text{adapt}} \cdot p_B(t),\ p_{\min},\ p_{\max}\Big) \tag{4}$$

where $k_{\text{adapt}} \in [0, 1]$ controls the sensitivity of the intervention to misinformation prevalence, and $p_{\min}$ and $p_{\max}$ denote the minimum and maximum allowable fact-checker densities, respectively, reflecting practical constraints on intervention resources. The clipping operator is defined as:

$$\text{clip}(x, a, b) = \min(\max(x, a), b) \tag{5}$$

which ensures that the adaptive control remains within feasible deployment limits.

This adaptive regulation mechanism allows the intervention intensity to scale with surges in misinformation prevalence. Conceptually, it parallels dynamic resource allocation strategies in epidemic control, where intervention strength increases during critical phases and decreases as the system stabilizes [11]. Importantly, the affine feedback structure provides interpretability, as the intervention level varies linearly with misinformation prevalence, while maintaining a clear separation between endogenous opinion dynamics and exogenous intervention control. The inclusion of clipping bounds prevents overreaction, numerical instability, and unrealistic deployment levels, thereby improving robustness under varying system conditions.

3.4. Spatially Informed Boundary Placement

Beyond adapting the quantity of fact-checkers through feedback control, our framework also dynamically adapts their spatial placement by targeting structurally influential nodes located at the interfaces between competing latent opinion clusters. These boundary nodes play a critical role in the propagation and suppression of misinformation, as they mediate direct interactions between fake-news (B) and non-fake-news (A) agents. Interventions at such interfaces can more effectively disrupt reinforcement mechanisms within misinformation clusters and promote opinion shifts through payoff-biased imitation.

Formally, a node v is defined as a boundary node if its neighbourhood contains at least one agent from each latent opinion class (A and B), excluding fact-checker nodes. Fact-checkers constitute an external intervention layer and are therefore not considered as part of the intrinsic opinion states for the purpose of boundary detection.

For each candidate node $v \in V$ we compute the normalized boundary interface score based on its local neighbourhood composition:

$$S_i^{\text{norm}} = \frac{n_B(i) \cdot n_{\neg B}(i)}{(k_i + \epsilon)^2} \tag{6}$$

where $n_B(v)$ is the number of neighbours whose latent opinion corresponds to fake news (B), $n_{\neg B}(v)$ is the number of neighbours with non-fake-news latent opinions (A), k_v is the degree of node v, and $\epsilon > 0$ is a small constant introduced for numerical stability. Fact-checker nodes are excluded from this calculation.

This normalized formulation was deliberately adopted to eliminate the degree bias inherent in the original multiplicative score. By dividing by the squared degree, the score reflects the relative composition of the neighbourhood (i.e., the product of the fractions of each opinion type) rather than its absolute count. Consequently, high-degree hubs are no longer disproportionately favoured unless they truly lie at the boundary between opposing opinion clusters. Nodes located deep inside homogeneous regions—whether entirely fake-news or truthful—receive a score of zero, ensuring that intervention resources are not wasted on non-boundary positions.

From a dynamic viewpoint, nodes with high boundary scores represent critical transition zones where opposing latent opinions directly interact. Placing a fact-checker at these locations alters local payoff structures, weakens the reinforcement of misinformation within echo chambers, and increases the probability of strategy adoption toward the truthful state via imitation dynamics. This design is theoretically supported by insights from complex contagion and percolation theory, which emphasise that boundary regions between competing domains often act as the primary drivers of large-scale transitions in networked systems [1,12].

At every T time step, the boundary interface scores are re-evaluated for all non-fact-checker nodes. Let

$$N_C(t) = \lfloor p_C(t) \cdot |V| \rfloor \tag{7}$$

denote the number of fact-checkers deployed at time t. The model then assigns fact-checker status to the $N_C(t)$ nodes with the highest boundary interface scores for the subsequent control interval.

Fact-checker allocation is implemented as a replacement-based intervention process: at each control update, the set of fact-checker nodes is fully redefined according to the current boundary scores and target density $p_C(t)$. Nodes that are no longer selected revert to their underlying latent opinions, potentially undergoing probabilistic post-intervention correction as described earlier. This design prevents uncontrolled accumulation of intervention resources and ensures that fact-checkers remain concentrated at structurally relevant interfaces as the system evolves. Between consecutive control updates, fact-checker assignments remain fixed, while agents continue to update their latent opinions according to the asynchronous imitation dynamics described in Section 3.2. Nodes assigned as fact-checkers are temporarily excluded from imitation dynamics and influence neighboring agents only through payoff interactions.

Overall, this spatially informed allocation strategy synthesizes principles from evolutionary game dynamics, complex contagion, and network control. Unlike static or randomly placed interventions, it continuously aligns resource deployment with both the current

system state and the evolving geometry of misinformation clusters, yielding a flexible and resilient mechanism for misinformation suppression. The resulting adaptive boundary-aware architecture is well suited for real-time deployment in algorithmic moderation systems, while remaining computationally efficient and robust under varying network conditions [17–20].

3.5. Analytical Approximation of Population Dynamics

To provide theoretical insight into the macroscopic behavior of the proposed framework, we derive a mean-field approximation describing the evolution of misinformation prevalence. The objective of this analysis is to capture dominant population-level trends induced by payoff-biased imitation and adaptive intervention, while abstracting from spatial correlations and network heterogeneity. Specifically, the approximation is constructed as a coarse-grained representation of the latent opinion dynamics (Section 3.2) and the adaptive intervention mechanism (Sections 3.3 and 3.4).

Let $\rho_A(t)$ and $\rho_B(t)$ denote the fractions of truthful and fake-news agents among opinion-bearing nodes, respectively, satisfying

$$\rho_A(t) + \rho_B(t) = 1 \tag{8}$$

Fact-checkers are excluded from this normalization, as they act as an external intervention layer rather than endogenous opinion states.

Under the mean-field assumption, agents interact with statistically representative neighbors. The temporal evolution of misinformation prevalence is approximated by the balance between imitation-driven growth $G(\rho_B)$ and intervention-induced suppression $L(\rho_B)$:

$$\frac{d\rho_B}{dt} = G(\rho_B) - L(\rho_B) \tag{9}$$

The growth term captures payoff-biased imitation and is modeled using a logistic-type form:

$$G(\rho_B) = \alpha \rho_B (1 - \rho_B) \tag{10}$$

where $\alpha > 0$ represents the effective propagation advantage of fake-news agents due to local reinforcement effects. The suppression term is assumed to be proportional to both misinformation prevalence and the density of fact-checkers:

$$L(\rho_B) = \gamma \, p_c(t) \, \rho_B \tag{11}$$

This formulation reflects the probabilistic influence of fact-checking interventions on agents' latent opinions, rather than deterministic state replacement. The parameter γ quantifies the effective strength of the suppression mechanism acting against misinformation. It captures the extent to which corrective interventions reduce the persistence and spread of misinformation within the population. In particular, γ governs the rate at which interactions with corrective influences translate into a net decline in misinformation density. Larger values of γ correspond to stronger suppression effects, leading to a faster erosion of misinformation clusters, whereas smaller values indicate weaker corrective influence and more persistent misinformation dynamics.

Substituting the adaptive regulation rule,

$$p_c(t) = p_{min} + k_{adapt} \rho_B(t) \tag{12}$$

the mean-field dynamics become

$$\frac{d\rho_B}{dt} = \alpha\rho_B(1-\rho_B) - \gamma\left(p_{min} + k_{adapt}\rho_B\right)\rho_B \tag{13}$$

Rearranging:

$$\frac{d\rho_B}{dt} = \rho_B\left[\alpha(1-\rho_B) - \gamma p_{min} - \gamma k_{adapt}\rho_B\right] \tag{14}$$

The system admits two equilibrium points: the misinformation-free equilibrium $\rho_B = 0$, and a nontrivial equilibrium

$$\rho_B^* = \frac{\alpha - \gamma p_{min}}{\alpha + \gamma k_{adapt}} \tag{15}$$

which exists when $\alpha > \gamma p_{\min}$.

Linearization around $\rho_B \approx 0$ yields $\frac{d\rho_B}{dt} \approx (\alpha - \gamma p_{\min})\rho_B$, indicating that the misinformation-free equilibrium is asymptotically stable when $\gamma p_{\min} > \alpha$.

This condition shows that a sufficiently strong baseline intervention guarantees extinction of misinformation at the population level. Moreover, increasing the adaptation gain k_{adapt} reduces the equilibrium prevalence ρ_B^*, analytically confirming the effectiveness of adaptive intervention in suppressing long-term misinformation. Although spatial boundary effects are not explicitly captured in this mean-field approximation, the model provides a consistent theoretical explanation for the simulation results presented in Section 4. In particular, the adaptive feedback introduces a density-dependent negative control term, which stabilizes the macroscopic dynamics and prevents large-scale misinformation outbreaks under varying system conditions.

3.6. Reproducibility and Implementation Details

All simulations were executed using an asynchronous update scheme. At each microscopic update step, a single non-fact-checker agent is selected uniformly at random, and its latent opinion is updated through payoff-biased imitation, following the pairwise comparison dynamics defined in Section 3.2. Agents assigned as fact-checkers are excluded from imitation dynamics during their assignment and influence neighboring agents solely through payoff interactions, without directly modifying latent opinions.

Adaptive intervention operates on a discrete-time control schedule. Fact-checker density and spatial placement are updated at fixed intervals of T time steps according to the feedback regulation and boundary-aware allocation mechanisms described in Sections 3.3 and 3.4. Between consecutive control updates, fact-checker assignments remain fixed while latent opinion dynamics continue to evolve asynchronously.

Network instances are generated prior to each simulation run, and agent opinions are initialized randomly, subject to specified initial conditions. All stochastic components of the model, including network generation, agent selection, and opinion updates, are driven by independent random seeds to ensure statistical validity. Each reported result is obtained by averaging over multiple independent realizations, with variability quantified using standard deviation, ensuring reproducibility and statistical robustness. This implementation protocol ensures that the model dynamics are fully specified and reproducible while maintaining a clear separation between the formal definition of the intervention mechanism and its empirical evaluation.

From an evolutionary perspective, the effectiveness of fact-checkers arises from their ability to reduce the relative fitness of fake-news strategies through payoff penalties applied during interactions. When the density of fact-checkers exceeds a critical level, the expected growth rate of misinformation becomes negative, leading to the gradual suppression

of misinformation clusters. This provides a qualitative threshold interpretation of the intervention mechanism, consistent with the mean-field analysis presented in Section 3.5.

3.7. Computational Complexity

The boundary detection procedure relies only on local neighborhood information. For a network with N nodes and average degree k, computing the interface score for all nodes requires O(Nk) operations, which scales linearly with network size for sparse graphs. Fact-checker allocation based on ranking the interface scores can be performed in O(NlogN) time using efficient sorting algorithms. Since both boundary detection and allocation rely on local or easily computable quantities, the overall intervention mechanism remains computationally efficient. Therefore, the proposed adaptive boundary-aware intervention framework is scalable and applicable to large-scale social networks under realistic computational constraints.

4. Experimental Evaluation and Results

4.1. Experimental Setup

To evaluate the proposed adaptive boundary-aware intervention model, we conducted agent-based simulations on networks with $N = 200$ nodes, consistent with prior work [3]. Three canonical network topologies were considered: small-world networks generated using the Watts–Strogatz model with rewiring probability $p = 0.1$ and average degree $k = 6$, scale-free networks generated using the Barabási–Albert model with $m = 3$, and random networks generated using the Erdős–Rényi model with $p = 0.03$ [21,22]. These topologies capture key structural properties of online social systems, including high clustering and short path lengths (small-world), hub-dominated connectivity (scale-free), and homogeneous random interactions (random networks). Considering all three ensures that the evaluation is not biased toward a specific structural regime and reflects diverse real-world interaction patterns.

All simulations were implemented in Python 3.11 using NetworkX (v3.6.1), NumPy (v2.4.1), SciPy (v1.17.0), pandas (v3.0.0) and matplotlib (v3.10.8). Each simulation was run for 2000 time steps and repeated over 30 independent random seeds to ensure statistical robustness and reduce sensitivity to stochastic initialization and network realization. The simulation horizon was selected to guarantee convergence to steady-state behavior under all tested configurations.

Two intervention strategies were evaluated. The first was a static baseline with fixed fact-checker density $p_C = 0.25$ and random placement. This baseline reflects common non-adaptive intervention strategies where resources are deployed uniformly without considering system state. The second was the proposed adaptive boundary-aware model, in which both the density $p_C(t)$ and spatial allocation of fact-checkers were dynamically updated based on real-time misinformation prevalence and the structural properties of the misinformation–truth interface.

Fact-checkers were implemented as an external overlay (C-set) on latent binary opinions (A/B). This separation allows the model to distinguish between temporary intervention effects and underlying belief states, which is essential for capturing realistic behavioral persistence. Nodes removed from the C-set retain their latent opinions but may undergo lasting correction with probability 0.7, modeling imperfect but persistent behavioral change.

The following metrics were recorded: (i) final fake-news count B, capturing long-term suppression performance; (ii) suppression time, measuring how quickly misinformation is controlled; (iii) system efficiency, capturing cumulative exposure; (iv) oscillation range, reflecting temporal stability; (v) echo-chamber statistics, capturing structural persistence; and (vi) mean fact-checker budget, quantifying resource usage.

System efficiency is defined as:

$$\text{Efficiency} = 1 - \frac{\text{AUC}_B}{\text{AUC}_{\max}} \tag{16}$$

where AUC_B is the area under $B(t)$, and $\text{AUC}_{\max}$ corresponds to the worst-case scenario with persistent misinformation. This metric captures both transient and steady-state performance, providing a more comprehensive evaluation than final-state metrics alone.

Unless otherwise stated, parameters were set to $\beta = 0.6$, $k_{\text{adapt}} = 0.7$, $T = 60$, $p_C^{\min} = 0.05$, and $p_C^{\max} = 0.50$, representing a balanced regime between responsiveness and stability.

4.2. Temporal Dynamics of Fake News Suppression

Figure 2 illustrates the temporal evolution of misinformation across network topologies. Without intervention, misinformation rapidly spreads and saturates the network, reaching a high-prevalence steady state. The static baseline reduces this prevalence but fails to eliminate misinformation, leaving a persistent residual population that reflects structural trapping within clusters.

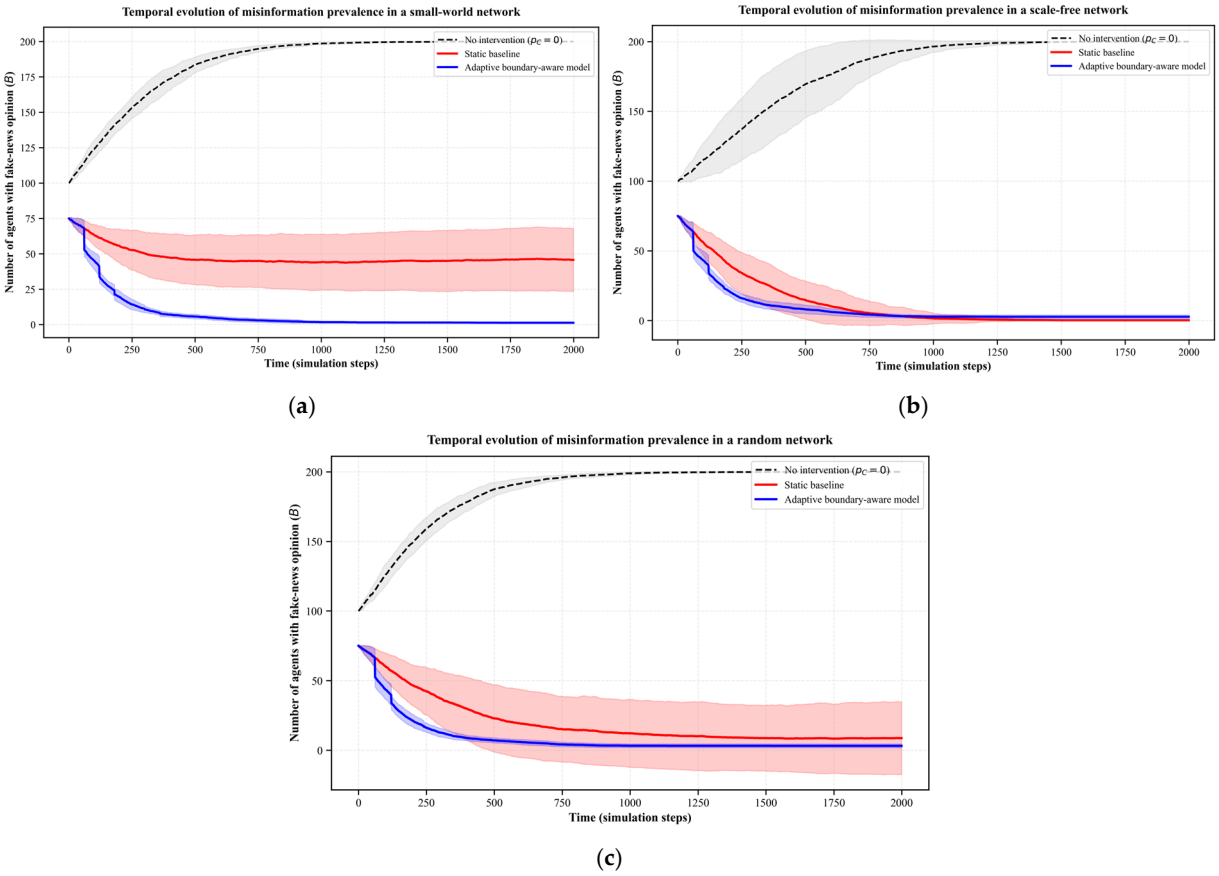


Figure 2. Temporal dynamics of fake-news suppression across (**a**) small-world, (**b**) scale-free, and (**c**) random networks.

In contrast, the adaptive model rapidly suppresses misinformation and converges to near-zero levels across all topologies. This improvement is driven by the coupled mechanism of feedback control and boundary targeting. When $B(t)$ is large, the controller increases intervention intensity, preventing uncontrolled growth. As misinformation clusters shrink, fact-checkers are reallocated toward boundary nodes—regions where fake-news and truthful agents interact—maximizing the probability of state transitions.

This dual adaptation in both quantity (density) and location (boundary targeting) allows the system to focus resources on high-impact regions, thereby accelerating suppression and reducing cumulative exposure. Importantly, the adaptive model reduces not only the final number of fake-news agents but also the total misinformation exposure over time, which is critical in real-world scenarios.

The shaded regions in Figure 2 indicate variability across runs. The adaptive model exhibits significantly narrower uncertainty bands, particularly after the transient phase, demonstrating improved stability and reduced sensitivity to stochastic fluctuations. This indicates that the model consistently converges to a low-prevalence equilibrium regardless of initial conditions.

Figure 3 shows the evolution of $p_C(t)$. The adaptive model initially increases intervention intensity to counteract rapid spread, then gradually reduces it as the system stabilizes. This results in a significantly lower average intervention budget (≈58.7% reduction) while maintaining superior suppression performance. This behavior reflects a self-regulating control mechanism that avoids over-allocation of resources.

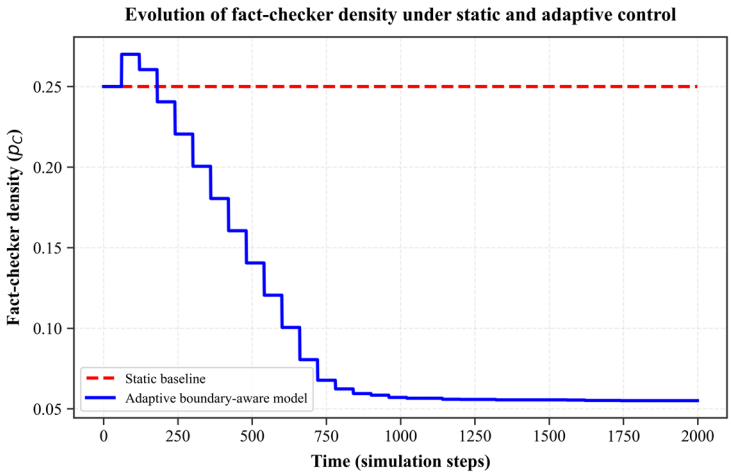


Figure 3. Evolution of fact-checker density under static and adaptive strategies.

4.3. Cross-Topology Robustness

Figure 4 compares final misinformation prevalence across network topologies. The adaptive model consistently achieves near-zero fake-news levels, while the baseline leaves a substantial residual population.

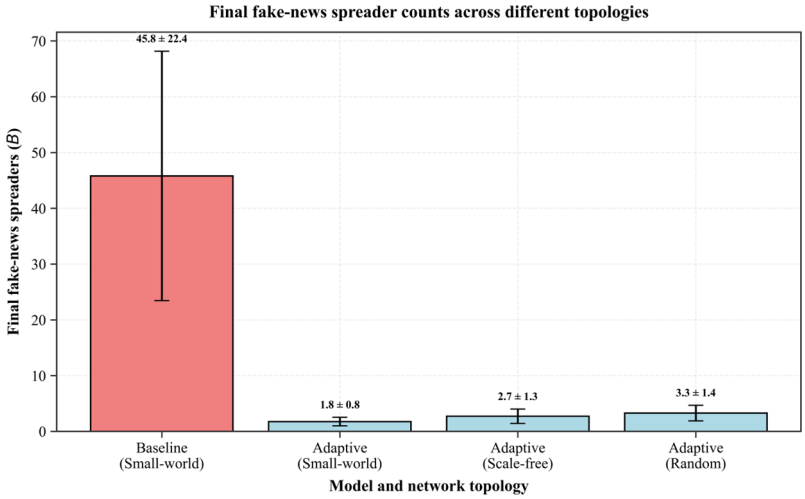


Figure 4. Final fake-news prevalence across different network topologies.

This robustness arises from the adaptability of boundary-aware placement to structural differences. In small-world networks, intervention targets clustered interfaces; in scale-free networks, it indirectly disrupts hub-driven cascades; and in random networks, it dissolves transient local clusters. This demonstrates that the proposed mechanism generalizes across heterogeneous network structures and is not topology-specific.

4.4. Quantitative Performance and Stability

Table 1 summarizes quantitative results. The adaptive model achieves a 97% reduction in final fake-news count and an 83.3% reduction in suppression time. Efficiency improves significantly, indicating reduced cumulative exposure.

Table 1. Quantitative comparison of suppression performance under static baseline and adaptive boundary-aware strategies.

Performance Metric	Baseline Model	Adaptive Model	Improvement
Final B Count	45.8 ± 22.4	1.37 ± 0.55	97.0% reduction
Suppression Time (steps)	1957.2 ± 230.3	327.2 ± 51.1	83.3% faster
Efficiency	0.762 ± 0.090	0.962 ± 0.004	26.2% increase
Oscillation Range	0.425 ± 0.283	0.000 ± 0.000	100% reduction
Number of B Echo Chambers	1.17 ± 0.52	0.00 ± 0.00	100% reduction
B in Chambers Ratio	0.958 ± 0.179	0.000 ± 0.000	100% reduction
Mean Fact-checker Budget	0.250 ± 0.000	0.103 ± 0.005	58.7% lower

Notably, oscillations are eliminated and echo chambers disappear entirely, indicating that the system converges to a stable and structurally clean state. Additionally, the mean fact-checker budget decreases substantially, demonstrating that improved performance is achieved alongside reduced resource consumption.

4.5. Statistical Significance and Distribution Analysis

To rigorously assess whether the observed performance improvements are statistically meaningful, we conducted an independent two-sample t-test on the final number of fake-news agents across simulation runs. The analysis yields a test statistic of $t = 10.70$ with $p < 0.001$, indicating that the difference between the baseline and adaptive models is highly statistically significant and cannot be attributed to random fluctuations.

Beyond statistical significance, it is equally important to evaluate the distributional characteristics of the outcomes. As illustrated in Figure 5, the baseline model exhibits a wide distribution with substantial variance, reflecting strong sensitivity to stochastic initial conditions and network realizations. In contrast, the adaptive model produces a sharply concentrated distribution near zero, with both lower mean and significantly reduced variance. This indicates that the proposed approach not only improves the expected performance but also enhances reliability and consistency across independent runs.

From a practical perspective, this reduction in variability is critical, as real-world social systems are inherently stochastic and uncertain. Therefore, an intervention strategy that consistently achieves low misinformation levels is more desirable than one that performs well only on average. Overall, the statistical and distributional analyses jointly confirm that the proposed adaptive boundary-aware model provides both significant and robust improvements over the static baseline.

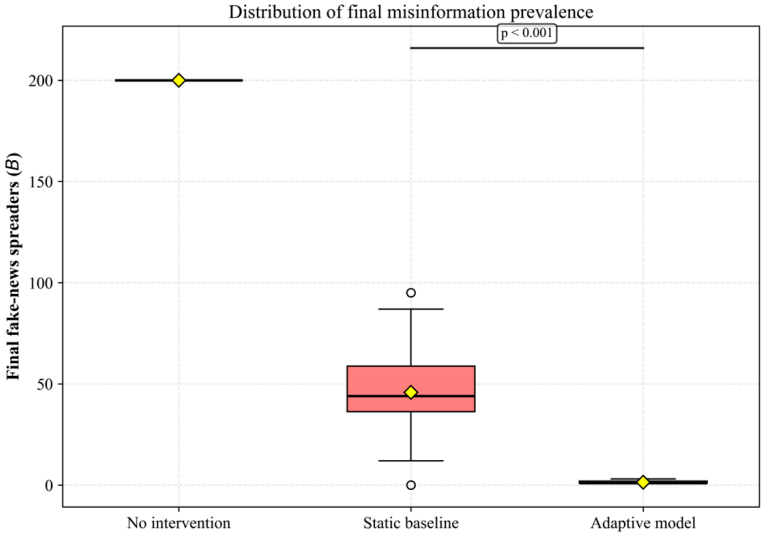


Figure 5. Distribution of final fake-news prevalence under different intervention strategies.

4.6. Parameter Sensitivity Analysis

To evaluate the robustness of the proposed model with respect to its key parameters, we performed a systematic sensitivity analysis on the selection intensity β, adaptation gain k_{adapt}, and control interval T. These parameters govern, respectively, the strength of behavioral imitation, the responsiveness of the adaptive controller, and the frequency of intervention updates.

The results (Figures 6–8) demonstrate that the model maintains strong performance across a broad parameter range. Increasing β slightly increases the residual misinformation level, reflecting stronger reinforcement of dominant strategies in the underlying evolutionary dynamics. However, the overall suppression performance remains near-optimal across all tested values, indicating that the model does not critically depend on a precise selection intensity.

Similarly, the adaptation gain k_{adapt} exhibits a non-monotonic effect: moderate values provide the best balance between responsiveness and stability, while excessively large values introduce mild fluctuations due to overreaction. Nevertheless, even in these regimes, the system remains stable and effective. The control interval T influences how frequently the intervention is updated; larger values reduce responsiveness but do not significantly degrade final outcomes, highlighting the robustness of the rate-limited control design.

Overall, these results indicate that the proposed framework operates reliably without requiring fine-tuned parameters, which is an important property for real-world deployment where exact parameter calibration is often infeasible.

To further assess the robustness of the proposed model with respect to behavioral assumptions, we conducted additional sensitivity analysis on payoff parameters, including the reward associated with fake-news interactions (B–B payoff) and the penalty imposed by fact-checkers (B–C payoff). Figure 9 illustrates the effect of varying these parameters on the final fake-news prevalence. The results show that the model maintains consistently low levels of misinformation across a broad range of payoff configurations. Although slight variations in final B values are observed, the overall suppression performance remains stable. This indicates that the effectiveness of the proposed intervention does not depend on a specific payoff setting and is robust under different incentive structures governing agent interactions.

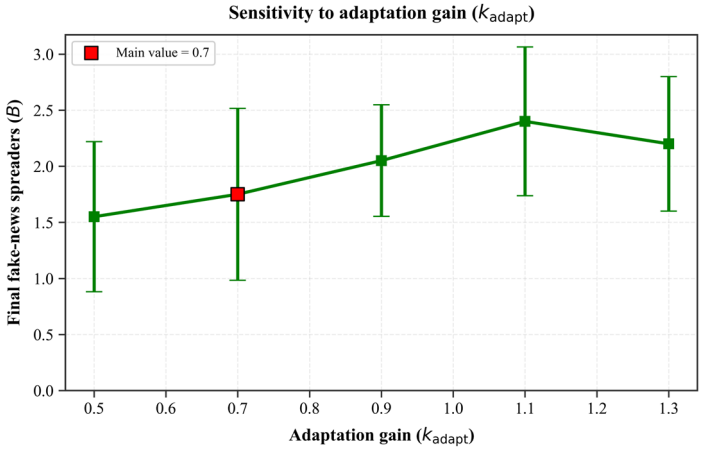


Figure 6. Sensitivity of misinformation suppression to adaptation gain k_{adapt}.

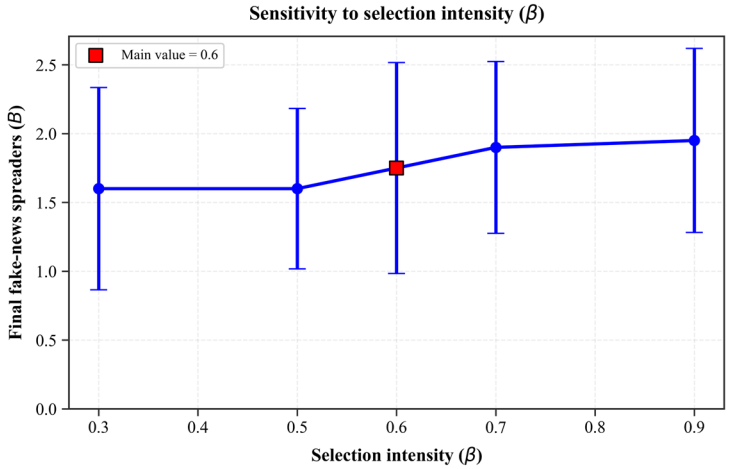


Figure 7. Sensitivity of misinformation suppression to selection intensity β.

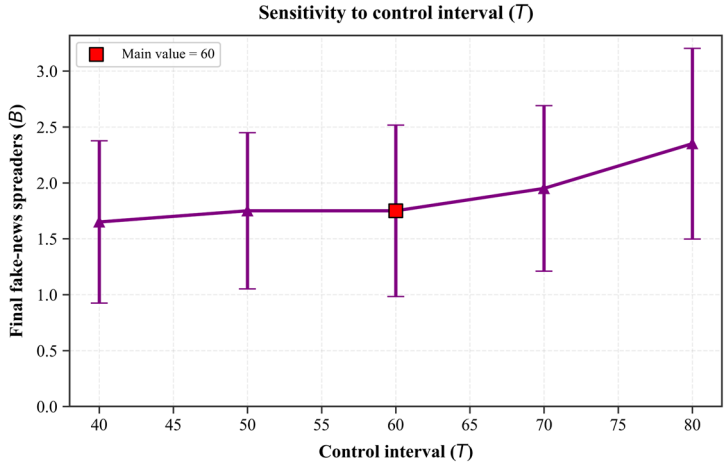


Figure 8. Sensitivity of misinformation suppression to selection interval T.

In addition, we evaluated the sensitivity of the model to the lasting correction probability, which determines the likelihood that a corrected agent permanently adopts truthful behavior after fact-checking intervention. As shown in Figure 10, the model remains robust across different probability values, with consistently low final misinformation levels. This result confirms that the performance of the proposed framework does not rely on fine-tuned behavioral assumptions regarding post-intervention persistence.

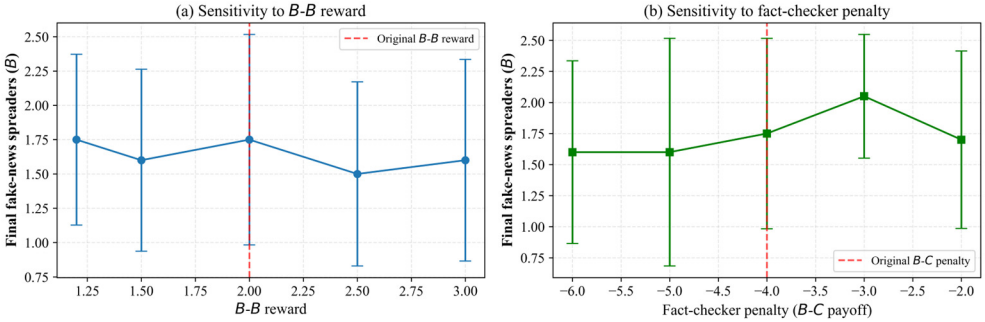


Figure 9. Sensitivity of misinformation suppression to payoff parameters: (**a**) variation in B–B interaction reward, and (**b**) variation in B–C penalty. The dashed vertical lines indicate the default parameter values used in the main experiments.

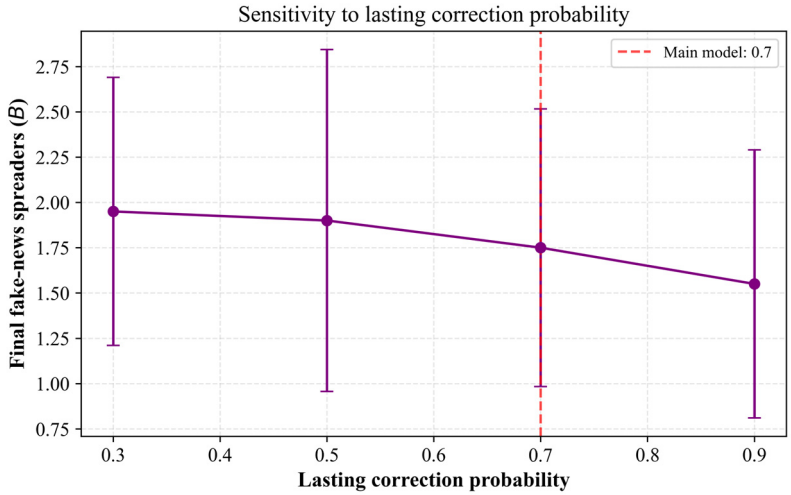


Figure 10. Sensitivity of misinformation suppression to the lasting correction probability. The dashed line indicates the default value used in the main model.

4.7. Ablation Analysis

To isolate the contribution of individual components of the proposed method, we conducted an ablation study in which adaptive density control and boundary-aware placement were evaluated independently and in combination. As shown in Figure 11, neither component alone is sufficient to achieve complete suppression of misinformation. Adaptive density without spatial targeting reduces overall prevalence but fails to eliminate localized clusters, while boundary-aware placement with fixed density improves targeting but lacks the responsiveness required to adapt to evolving system states.

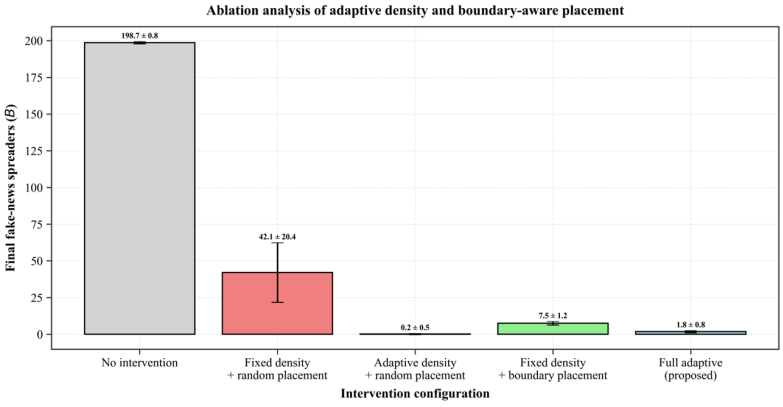


Figure 11. Ablation analysis of intervention components.

Only the full model—combining adaptive density control with boundary-aware placement—achieves near-complete suppression across all runs. This demonstrates that the effectiveness of the proposed approach arises from the synergistic interaction between temporal adaptation and structural targeting, rather than from either mechanism in isolation. In particular, dynamic adjustment of intervention intensity ensures timely response to global system changes, while boundary-aware placement ensures that resources are deployed at structurally critical locations.

4.8. Structural Effects: Echo Chambers

In addition to reducing overall misinformation prevalence, the proposed model significantly alters the structural organization of the network. Under the baseline strategy, misinformation tends to persist in the form of echo chambers—densely connected clusters of agents sharing the same false belief. These structures act as reservoirs that sustain misinformation even when global prevalence decreases.

In contrast, the adaptive boundary-aware model eliminates both the number of echo chambers and the fraction of fake-news agents contained within them. This indicates that the proposed intervention disrupts not only the diffusion process but also the structural persistence mechanisms that stabilize misinformation communities. By targeting boundary nodes, the model weakens the cohesion of these clusters and prevents their long-term survival, leading to a structurally cleaner and more stable equilibrium.

These findings are further supported by a structural analysis of the network, as shown in Figure 12. Under the static baseline, misinformation agents form persistent and highly concentrated echo chambers, where a large fraction of B nodes remain isolated from corrective influence. In contrast, the adaptive model eliminates both the number of echo chambers and the concentration of misinformation within them. This demonstrates that the proposed intervention not only reduces overall misinformation prevalence but also disrupts the structural mechanisms that sustain it, leading to complete fragmentation of misinformation clusters.

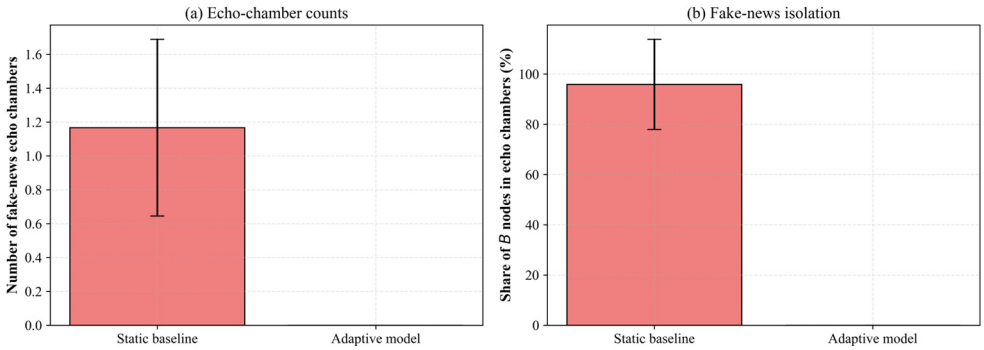


Figure 12. Structural analysis of misinformation echo chambers: (**a**) number of fake-news echo chambers, and (**b**) proportion of fake-news agents located within such chambers.

4.9. Robustness Under Noisy Observation

To evaluate robustness under imperfect information, we introduced observation noise ϵ, which perturbs the perceived states of neighboring agents. Figure 13 shows that the adaptive model maintains near-zero misinformation levels even as noise increases, whereas the baseline remains largely unaffected at a high misinformation level.

This robustness can be explained by the feedback-driven nature of the adaptive controller. While local observations may be noisy, the global feedback mechanism integrates information over the network and compensates for local errors through aggregate adjustment of $p_C(t)$. As a result, the system remains stable and continues to suppress

misinformation effectively even under uncertainty. This property is particularly important for real-world applications, where perfect information is rarely available.

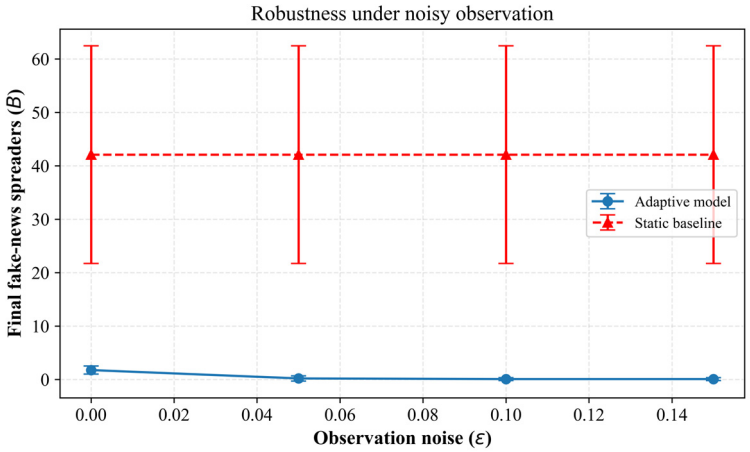


Figure 13. Robustness of the adaptive model under noisy observation.

4.10. Boundary Score Validation

The effectiveness of the proposed intervention critically depends on accurately identifying boundary nodes, defined as nodes whose neighborhoods contain a mixture of truthful (A) and misinformation (B) agents. These nodes represent structurally sensitive regions where opinion transitions are most likely to occur and, therefore, where intervention is most impactful.

As defined in Section 3.4, our method employs a normalized boundary score that eliminates degree bias by dividing the product of mixed neighbors by the squared degree. For comparison, we also evaluate the unnormalized (original) formulation:

$$S_i^{\text{orig}} = n_B(i) \cdot n_{\neg B}(i) \tag{17}$$

where $n_B(i)$ and $n_{\neg B}(i)$ denote the number of neighbors holding fake-news and non-fake-news opinions, respectively. While this unnormalized formulation captures the presence of mixed neighborhoods, it implicitly scales with node degree, leading to disproportionately large scores for high-degree nodes. As a result, intervention resources may be biased toward hubs even when they are not located at critical transition boundaries.

As shown in Figure 14, the normalized boundary score (defined in Section 3.4) leads to consistently lower final misinformation prevalence and reduced variability across simulation runs compared to the unnormalized formulation (Equation (17)). This result indicates that degree-aware normalization improves the precision of boundary detection and enhances the effectiveness of intervention allocation. More broadly, it highlights the importance of properly scaling structural indicators in network-based control strategies, particularly in heterogeneous topologies where degree distributions are highly skewed.

4.11. Scalability Analysis

To assess scalability, we evaluated the model on networks of increasing size. The relative improvement is defined as:

$$\text{Improvement} = \frac{B_{\text{baseline}} - B_{\text{adaptive}}}{B_{\text{baseline}}} \times 100. \tag{18}$$

As shown in Figure 15, the relative improvement decreases gradually with network size, which can be attributed to increased structural complexity and stronger persistence of

misinformation clusters in larger networks. Nevertheless, the adaptive model continues to provide substantial suppression even at larger scales.

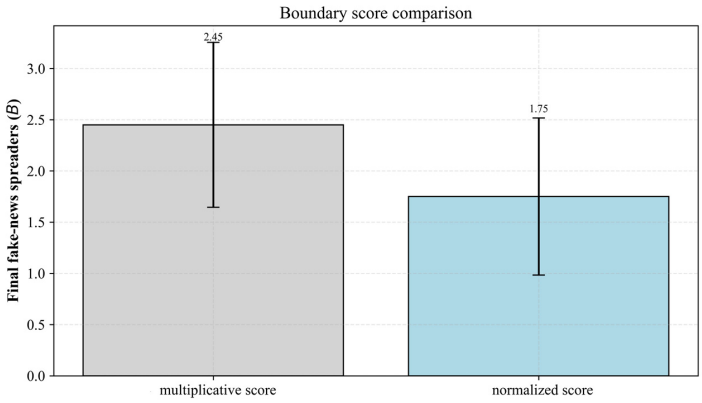


Figure 14. Comparison of original and normalized boundary scoring methods, showing improved suppression performance under the normalized formulation.

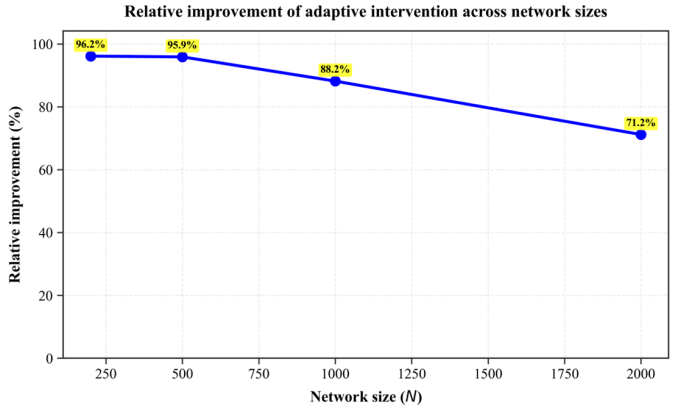


Figure 15. Relative improvement across network sizes.

Complementary to this relative measure, Figure 16 presents the absolute number of remaining fake-news agents under both strategies. The results indicate that the absolute reduction in misinformation increases significantly with network size, demonstrating that the practical impact of the intervention becomes more pronounced in larger systems. This behavior highlights an important distinction: although larger systems are inherently more difficult to fully control, the adaptive strategy remains effective in reducing the total number of affected agents. Therefore, the proposed model exhibits strong scalability and remains applicable to large-scale social networks.

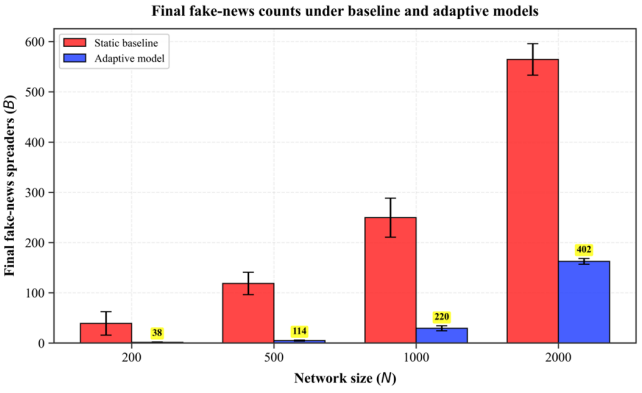


Figure 16. Absolute misinformation counts across different network sizes.

5. Discussion

A comprehensive evaluation across multiple dimensions provides strong evidence for the effectiveness and robustness of the proposed adaptive boundary-aware intervention framework in mitigating misinformation spread in complex networks. Compared with the static baseline model by Jones et al. [3], the proposed method achieves substantial improvements, including approximately a 97% reduction in final fake-news prevalence, more than 80% acceleration in suppression time, and about a 26% increase in overall system efficiency. These gains are not only statistically significant but also practically meaningful, particularly in scenarios where rapid containment and efficient use of limited intervention resources are critical. The observed performance improvements arise from the integration of two complementary mechanisms. First, the feedback-driven adaptation of fact-checker density $p_C(t)$ enables dynamic adjustment of intervention intensity in response to real-time misinformation prevalence. This aligns the model with principles of adaptive control under uncertainty, allowing the system to respond to evolving diffusion dynamics rather than relying on fixed intervention levels. Second, the boundary-aware spatial allocation strategy prioritizes nodes located at the interfaces between misinformation and truthful regions. These boundary nodes represent structurally critical transition points where opinion changes are most likely to occur. By concentrating resources at these interfaces, the model effectively disrupts the propagation pathways of misinformation at their most sensitive locations. The joint operation of these two mechanisms allows simultaneous control over both the scale and spatial distribution of interventions, resulting in significantly improved suppression efficiency compared to static or random allocation strategies. The results further demonstrate that the proposed framework achieves not only stronger suppression but also stable convergence behavior. In particular, the elimination of oscillations and the disappearance of echo chambers indicate that the intervention does not merely reduce misinformation temporarily, but also removes the structural conditions that sustain it. This structural effect is especially important in networked systems, where clustered misinformation can persist even under moderate intervention. The ability of the proposed method to dissolve such clusters suggests that targeting boundary regions provides a powerful mechanism for destabilizing and eliminating misinformation at both local and global scales. This directly addresses concerns regarding the structural validity of the boundary-based intervention mechanism.

Importantly, the framework exhibits strong robustness across different network topologies, including small-world, scale-free, and random networks. This indicates that the effectiveness of the approach does not depend on specific structural assumptions. In small-world networks, the method targets clustered interfaces; in scale-free networks, it indirectly disrupts hub-mediated diffusion; and in random networks, it rapidly dissolves weakly structured misinformation regions. This consistent performance across heterogeneous structures highlights the generalizability of the proposed approach. These findings further reinforce that the model is not overfitted to a specific topology and remains valid under diverse structural conditions.

Additional sensitivity analyses further confirm the robustness of the model. The results show that performance remains stable across a wide range of selection intensities, adaptation gains, and update intervals, indicating that the method does not rely on finely tuned parameters. Moreover, extended experiments on payoff parameters, including the reward associated with misinformation interactions and the penalty imposed by fact-checkers, demonstrate that the model maintains strong suppression performance under varying behavioral incentives. This directly addresses concerns regarding sensitivity to payoff structures and confirms that the proposed mechanism remains effective under different behavioral assumptions.

From a practical perspective, an important consideration is the cost associated with frequent reallocation of fact-checkers. In the current model, fact-checker placement is updated periodically to reflect changes in network state. To assess the impact of such operational constraints, we introduced an extension incorporating placement inertia, in which a fraction of fact-checkers retain their previous assignments across update intervals. The results indicate that the model remains robust across a broad range of inertia levels, with only minor performance degradation, thereby addressing concerns related to switching costs and implementation feasibility.

Another important aspect is the reliability of the feedback signal used for adaptive control. In real-world scenarios, misinformation prevalence may be observed with noise or delay. Our results under noisy observation conditions show that the feedback-driven mechanism remains stable and effective even when the observed state is imperfect. This robustness arises from the aggregate nature of the control signal, which mitigates the impact of local observation errors and directly addresses concerns regarding noise sensitivity in real-world deployments.

The scalability analysis further supports the practical applicability of the proposed framework. Although relative performance gains decrease as network size increases—due to the stronger persistence of large misinformation clusters—the absolute reduction in misinformation grows substantially. This indicates that the method remains effective in large-scale systems and that its practical impact becomes even more significant as the size of the network increases. This directly responds to concerns regarding the applicability of the model beyond small-scale simulations.

Despite these strengths, several limitations should be acknowledged. First, the current model assumes instantaneous reassignment of intervention resources within each control interval. While the placement inertia extension partially addresses this issue, more detailed modeling of switching costs and delayed responses could further improve realism. Second, the intervention strategy primarily focuses on boundary targeting and does not explicitly incorporate deep penetration into the core of highly cohesive misinformation clusters. Future work could explore hybrid strategies that combine boundary targeting with diffusion-based or multi-hop intervention mechanisms.

Furthermore, aggressive penalties may in some cases trigger a backfire effect, whereby individuals strengthen their commitment to misinformation due to psychological reactance. Such behavioral responses are not captured in the current payoff-based imitation model and warrant investigation in future extensions.

Third, while the mean-field approximation provides useful analytical insights, it remains a coarse-grained representation of network dynamics, and more refined theoretical models could further enhance understanding of the observed phenomena. This clarification directly addresses concerns regarding the limitations of mean-field approximations.

Overall, the findings highlight the critical role of structural awareness in misinformation mitigation. Rather than relying solely on increasing intervention intensity, the results demonstrate that strategically allocating resources to structurally sensitive regions—combined with adaptive feedback control—provides a more effective and efficient approach to suppressing misinformation in complex networks.

6. Conclusions

This study introduces an adaptive boundary-aware intervention framework for controlling the spread of fake news in complex social networks, in which both the density and spatial allocation of fact-checkers are dynamically adjusted based on evolving system conditions. The proposed model integrates feedback-based regulation of intervention intensity with structure-aware placement, enabling simultaneous control over when and

where interventions are applied. By modeling fact-checkers as an external intervention layer that does not participate in imitation dynamics, the framework distinguishes between endogenous opinion evolution and exogenous control actions, thereby providing a more realistic representation of moderation mechanisms. Furthermore, the incorporation of a probabilistic lasting correction mechanism captures the non-deterministic but persistent influence of fact-checking on agents' latent opinions, improving the behavioral realism of the model. Extensive simulations demonstrate that this integrated approach significantly outperforms static strategies, achieving near-complete suppression of misinformation, faster convergence, and improved efficiency, while requiring fewer intervention resources. The framework also shows strong robustness across parameter settings and network topologies, as well as favorable scalability properties as network size increases. From a broader perspective, the results highlight the importance of combining adaptive control with structural awareness in designing effective intervention strategies for complex networked systems. Although the current model assumes perfect detection of misinformation and idealized intervention mechanisms, future work will extend the framework by incorporating imperfect detection, learning-based classification, delayed interventions, and validation on real-world network data. Overall, the proposed approach provides a scalable, efficient, and theoretically grounded foundation for misinformation mitigation in large-scale socio-technical systems.

Author Contributions: Conceptualization, M.T.F., G.N. and N.M.; Methodology, M.T.F. and G.N.; Software, M.T.F. and G.N.; Validation, M.T.F., R.G. and N.M.; Formal Analysis, M.T.F., R.G. and N.M.; Investigation, M.T.F., R.G. and N.M.; Writing—Original Draft Preparation, G.N.; Writing—Review and Editing, M.T.F., R.G. and N.M.; Visualization, M.T.F. and G.N.; Supervision, M.T.F., R.G. and N.M.; Project Administration, N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT, 5.3 for the purposes of English editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science* **2018**, *359*, 1146–1151. [CrossRef] [PubMed]
2. Del Vicario, M.; Bessi, A.; Zollo, F.; Petroni, F.; Scala, A.; Caldarelli, G.; Stanley, H.E.; Quattrociocchi, W. The Spreading of Misinformation Online. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 554–559. [CrossRef] [PubMed]
3. Jones, M.I.; Pauls, S.D.; Fu, F. Containing Misinformation: Modeling Spatial Games of Fake News. *PNAS Nexus* **2024**, *3*, pgae090. [CrossRef] [PubMed]
4. Törnberg, P. Echo Chambers and Viral Misinformation: Modeling Fake News as Complex Contagion. *PLoS ONE* **2018**, *13*, e0203958. [CrossRef] [PubMed]
5. Shao, C. The Spread of Low-Credibility Content by Social Bots. *Nat. Commun.* **2018**, *9*, 4787. [CrossRef] [PubMed]
6. Hsu, C.-C.; Ajorlou, A.; Jadbabaie, A. A Game-Theoretic Model of Misinformation Spread in Social Networks. *Games Econ. Behav.* **2025**, *153*, 386–407. [CrossRef]
7. Meng, Y.; Broom, M.; Li, A. Impact of Misinformation in the Evolution of Collective Cooperation on Networks. *J. R. Soc. Interface* **2023**, *20*, 20230295. [CrossRef] [PubMed]

8. Liu, J.; Song, M.; Fu, G. Intervention Analysis for Fake News Diffusion: An Evolutionary Game Theory Perspective. *Nonlinear Dyn.* **2024**, *112*, 14657–14675. [CrossRef]
9. Kempe, D.; Kleinberg, J.; Tardos, É. Maximizing the Spread of Influence through a Social Network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 137–146.
10. Budak, C.; Agrawal, D.; El Abbadi, A. Limiting the Spread of Misinformation in Social Networks. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 665–674.
11. Govindankutty, S.; Gopalan, S.P. Epidemic Modeling for Misinformation Spread in Digital Networks through a Social Intelligence Approach. *Sci. Rep.* **2024**, *14*, 19100. [CrossRef] [PubMed]
12. Acemoglu, D.; Ozdaglar, A.; Siderius, J. A Model of Online Misinformation. *Rev. Econ. Stud.* **2024**, *91*, 3117–3150. [CrossRef]
13. Jing, H. Optimal Control of Rumor Propagation Model Considering Individual Activity and Platform Effect. In Proceedings of the 8th International Symposium on Big Data and Applied Statistics (ISBDAS), Guangzhou, China, 28 February–2 March 2025; pp. 729–733.
14. Jiang, J.; Chen, X.; Huang, Z.; Li, X.; Du, Y. Deep Reinforcement Learning-Based Approach for Rumor Influence Minimization in Social Networks. *Appl. Intell.* **2023**, *53*, 20293–20310. [CrossRef] [PubMed]
15. Shahi, S.; Dirkson, A.; Majchrzak, T. An Exploratory Study of COVID-19 Misinformation on Twitter. *Online Soc. Netw. Media* **2021**, *22*, 100104. [CrossRef] [PubMed]
16. Phan, X.; Guo, Q.; Zubiaga, A. Fake News Detection: A Survey of Graph Neural Network Methods. *Appl. Soft Comput.* **2023**, *139*, 110235. [CrossRef] [PubMed]
17. Ma, L.; Li, B.; Wang, J. Dynamic Evolutionary Analysis of Opinion Leaders' and Netizens' Uncertain Information Dissemination Behavior Considering Random Interference. *Front. Phys.* **2024**, *12*, 1387312. [CrossRef]
18. Nowak, M.A. *Evolutionary Dynamics: Exploring the Equations of Life*; Harvard University Press: Cambridge, MA, USA, 2006.
19. Rivera, D.E.; Pew, M.D.; Collins, L.M. Using Engineering Control Principles to Inform the Design of Adaptive Interventions: A Conceptual Introduction. *Drug Alcohol Depend.* **2007**, *88*, S11–S17. [CrossRef] [PubMed]
20. Yu, W.; Chen, G.; Lü, J. On Pinning Synchronization of Complex Dynamical Networks. *Automatica* **2009**, *45*, 429–435. [CrossRef]
21. Watts, D.J.; Strogatz, S.H. Collective Dynamics of "Small-World" Networks. *Nature* **1998**, *393*, 440–442. [CrossRef] [PubMed]
22. Barabási, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.