A Computational Study of Misinformation Diffusion and Public Opinion Formation in Online News Networks

Shane Makinen
Department of Computer Science, University of North Texas, Denton, TX, USA.
shanemakinen68@unt.edu

Porge Anderson
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
janderson@colostate.edu

Leixin Jia
Department of Computer Science, University of Houston, Houston, TX, USA.
leixinwork@uh.edu

Abstract

The proliferation of misinformation within online news networks presents a critical challenge to democratic discourse and public trust. This paper presents a computational investigation into the mechanisms by which false or misleading information spreads across digital platforms and shapes collective opinion. We develop an agent-based modeling framework that integrates network topology, cognitive biases, and platform-driven content curation algorithms to simulate the coupled dynamics of information diffusion and opinion formation. The study emphasizes system-level architectural choices and their structural trade-offs, examining how different governance strategies—including content moderation, algorithmic transparency, and user empowerment—affect the robustness and fairness of information ecosystems. Our simulations reveal that homophilic network structures and reinforcement learning in recommender systems amplify echo chambers and accelerate the spread of low-veracity content, while diversity-inducing algorithms can increase resilience but may incur costs in user engagement and personalization accuracy. We further analyze the sustainability of intervention policies under adversarial manipulation and assess the implications for socio-technical infrastructure design. The findings underscore the need for multi-stakeholder governance approaches that balance freedom of expression with collective epistemic welfare, and highlight critical directions for future research in computational social science, human-computer interaction, and public policy.

Keywords

misinformation diffusion, public opinion formation, online news networks, computational modeling, algorithmic governance, robustness, fairness, socio-technical infrastructure.

1. Introduction

The digital transformation of news consumption has fundamentally altered how individuals access, evaluate, and share information. Online platforms aggregate content from diverse sources and deploy algorithmic curation systems to personalize user feeds, thereby shaping exposure to both credible journalism and fabricated narratives. The resulting information networks exhibit complex properties that cannot be understood solely through linear causal models; instead, they demand a systems-oriented approach that captures feedback loops between individual behavior, platform design, and collective outcomes. Misinformation—defined as false or misleading content produced without malicious intent, as distinct from disinformation—poses a systemic risk because it exploits cognitive vulnerabilities and network amplification mechanisms to erode the epistemic foundations of public debate [1, 2]. Understanding the underlying computational processes is essential for designing interventions that preserve the integrity of democratic deliberation without undermining free expression.

Early research on misinformation diffusion relied primarily on observational studies of social media traces, revealing that false news spreads significantly faster, farther, and more broadly than true news on platforms such as Twitter and Facebook [1]. Subsequent work identified individual-level factors, such as analytical thinking and political ideology, that modulate susceptibility to misinformation [2, 3]. However, these studies often treat the platform as a passive conduit rather than an active participant in shaping diffusion dynamics. More recent computational approaches have begun to model the co-evolution of content propagation and opinion evolution within structured populations, incorporating elements from network science, statistical physics, and multi-agent simulation [4, 5]. Yet many such models remain stylized and abstract, limiting their direct applicability to real-world governance questions.

This paper advances a comprehensive computational framework that integrates three critical dimensions: the topological structure of online news networks, the behavioral rules governing user attention and belief updating, and the algorithmic mediation of content exposure. We adopt a system-level perspective that foregrounds architectural trade-offs—for example, between maximization of user engagement and minimization of misinformation spread—and evaluates the implications for infrastructure sustainability, robustness against adversarial attacks, and fairness across demographic groups. By simulating a range of policy scenarios, including fact-checking integration, content moderation thresholds, and algorithmic diversity injection, we provide evidence for the conditions under which interventions succeed or fail. The remainder of the paper is organized as follows. Section 2 reviews related work and situates our contribution within the broader literature. Section 3 describes the modeling framework and simulation design. Section 4 presents the analysis of structural dynamics, including echo chamber formation and cascade patterns. Section 5 discusses governance and policy implications, with attention to fairness and robustness. Section 6 concludes with a synthesis of findings and directions for future research.

2. Background and Related Work

The study of misinformation diffusion sits at the intersection of computational social science, network science, and human-computer interaction. Early foundational work in social network analysis established that information spreads through ties of friendship and affiliation following patterns of simple and complex contagion [6]. Subsequent research demonstrated that the structure of these networks—particularly the degree of homophily and the density of cross-cutting ties—strongly influences the speed and breadth of information cascades [7, 8]. In the context of online news, homophilic communities tend to reinforce shared beliefs and filter out dissonant information, creating echo chambers that facilitate the uncritical propagation of false claims [4, 5]. Computational models have shown that even a small fraction of highly connected nodes can dramatically accelerate the reach of a rumor, especially when those nodes are automated accounts or coordinated groups [9].

Algorithmic curation represents a second major driver of misinformation dynamics. Recommender systems that optimize for click-through rates or time spent on platform inadvertently prioritize sensational or emotionally charged content, which often aligns with low-veracity information [2, 10]. The feedback loop between user engagement and content promotion creates a perverse incentive structure: content that provokes strong reactions is more likely to be amplified, regardless of its factual accuracy. Several studies have documented the role of social bots in magnifying false narratives, particularly during elections and public health crises [11]. Meanwhile, interventions such as fact-checking tags or source credibility labels have shown mixed effectiveness, partly because users interpret them through preexisting ideological lenses [3]. These findings underscore the need for models that capture the interplay between human cognition and algorithmic mediation.

From a governance perspective, the literature has explored a spectrum of policy approaches, ranging from platform self-regulation and transparency mandates to state-sponsored content filtering and digital literacy campaigns [12]. Each approach entails trade-offs: heavy-handed moderation risks censorship and backlash, while purely laissez-faire regimes allow harmful content to run rampant. A robust framework must account for the dynamic nature of misinformation, which evolves in response to countermeasures. Adversarial agents can shift strategies—for example, by using coded language or reposting content across platforms—to evade detection [13]. Thus, any sustainable governance solution must be adaptable, multi-layered, and informed by continuous computational monitoring. The present study contributes to this conversation by providing a quantitative evaluation of several intervention architectures within a unified simulation environment.

3. Modeling Framework and System Architecture

We design an agent-based computational model that simulates a population of users connected via a directed social network representing their subscriptions or friendship ties on a news platform. Each agent possesses a latent belief about a set of topics, operationalized as a continuous variable that updates when exposed to new information. The information environment consists of news items assigned a ground-truth veracity value, which may be high (accurate) or low (misinformation). Agents probabilistically share items with their neighbors, with sharing probability modulated by the item's emotional valence, alignment with their existing beliefs (confirmation bias), and the credibility of the source. The platform's recommender system surfaces content to users based on predicted engagement, computed using a reinforcement learning algorithm that balances exploration and exploitation. This architecture captures the core feedback loop: user interactions influence the recommender, which in turn shapes future exposure.

To represent the network structure, we generate synthetic graphs with varying levels of homophily and modularity, parameterized by a degree distribution that follows a power law, consistent with empirical observations of social networks [7]. Agents are assigned to latent opinion groups, and the probability of a connection between two agents decreases with opinion distance, mimicking homophilic tie formation. The recommender system is modeled as a multi-armed bandit, where each arm corresponds to a content category or source; the system learns to favor arms that historically yielded high engagement, but with a tunable exploration parameter. We introduce stochasticity to account for algorithmic noise and externalities such as trending topics or news cycles.

The simulation proceeds in discrete time steps. At each step, the platform selects a subset of items to display to each agent based on the recommender's current policy. Agents observe these items, update their beliefs according to a Bayesian-like rule adjusted for confirmation bias, and decide whether to reshare. Shared items enter a global feed that competitors the recommender system. We calibrate model parameters using empirical findings from observational studies: for example, the probability of sharing false news is set higher than true news under conditions of high emotional arousal [1], and the confirmation bias strength is drawn from distributions reported in cognitive psychology experiments [2]. The simulator runs for hundreds of time steps per scenario, and outcomes are averaged over multiple independent runs to ensure statistical reliability.

This architecture enables systematic exploration of design parameters that correspond to real-world policy levers. For instance, we can vary the recommender's exploration rate to increase or decrease the diversity of content shown, simulate the effect of inserting fact-checking tags that reduce the believability of flagged items, or model the impact of removing accounts identified as coordinated misinformation sources. By comparing the resulting diffusion curves, opinion polarization indices, and network-level measures of information quality, we derive insights into the structural trade-offs inherent in different governance regimes.

4. Structural Dynamics of Information Spread

Our simulations reveal several robust patterns concerning the interplay between network topology, algorithmic curation, and misinformation diffusion. First, networks with high homophily exhibit pronounced echo chamber formation: agents overwhelmingly interact with like-minded peers, causing their beliefs to converge at extremes and rendering them resistant to corrective information. Under such conditions, false news cascades achieve nearly complete penetration within opinion clusters before any cross-cutting exposure occurs, consistent with prior theoretical predictions [4, 8]. The recommender system exacerbates this effect when it optimizes for engagement, because the most engaging content within a homophilic cluster is often the most polarizing or sensationalist. Consequently, the platform inadvertently entrenches existing cleavages rather than bridging them.

Second, introducing algorithmic diversity—for example, by forcing the recommender to reserve a fraction of exposure for content from outside an agent's usual neighborhood—lowers the peak reach of misinformation cascades and reduces belief polarization. However, this intervention comes at a cost: user engagement metrics decline by approximately fifteen to thirty percent relative to the baseline, depending on the diversity parameter, because agents encounter less tailored content. This trade-off between epistemic accuracy and user satisfaction is a central design challenge for platform architects. Our analysis suggests that the decline in engagement is not uniform across user groups; agents with moderate initial beliefs benefit more from diverse exposure, while highly partisan agents may disengage entirely or migrate to alternative platforms, a phenomenon known as platform exit. Thus, policymakers must consider equity implications when designing diversity mandates.

Third, we examine the robustness of the information ecosystem to adversarial manipulation. We simulate a scenario where a small fraction of agents are controlled by a malicious entity that strategically injects and amplifies false narratives. In the baseline architecture, these adversarial agents can trigger cascades that affect up to forty percent of the population before detection. When we incorporate a simple bot-detection algorithm that identifies accounts with abnormally high sharing rates and low out-degree reciprocity, cascade size reduces by roughly half. However, the detection algorithm also produces false positives, leading to the removal of legitimate users and raising fairness concerns. The trade-off between sensitivity and specificity is acute; overly aggressive filtering can suppress minority voices and reduce overall network resilience [14]. This finding highlights the need for transparency and appeals mechanisms in automated content governance systems.

Fourth, we explore the effect of fact-checking interventions deployed either pro-actively (by tagging content before it spreads widely) or reactively (after a cascade has begun). Proactive tagging, though more resource-intensive, prevents large portions of the cascade from ever reaching susceptible users. Reactive tagging has a more modest impact, mainly slowing rather than halting spread, and sometimes even backfiring by drawing attention to the contested content—a phenomenon observed in real-world studies [15]. The simulations indicate that the effectiveness of fact-checking depends critically on the trustworthiness of the fact-checking source in the eyes of the audience. If the fact-checker is perceived as biased against a political group, exposure to its tags can strengthen rather than weaken belief in the misinformation among that group, a form of motivated reasoning. Therefore, the design of fact-checking infrastructure must be complemented by efforts to build cross-partisan credibility.

Finally, we consider the sustainability of these interventions over long time horizons. As the simulation runs for several hundred time steps, we observe that without continuous adaptation, misinformation sources evolve to evade detection—for example, by altering the linguistic framing of false claims or rotating accounts. A static detection model degrades in performance, while an adaptive model that retrains periodically maintains higher accuracy. This finding underscores the need for platforms to invest in ongoing learning and updating of their governance algorithms, a requirement that imposes computational and economic costs [16]. Sustainability thus becomes a resource allocation problem: platforms must balance the benefits of reduced misinformation against the expenses of system maintenance and the potential for overfitting to current adversary strategies.

5. Robustness, Fairness, and Policy Implications

The simulation results carry important implications for the design of robust and fair socio-technical infrastructures for online news. Robustness, in this context, refers to the ability of the information ecosystem to maintain high average veracity and low polarization despite perturbations such as coordinated disinformation campaigns, sudden shifts in public attention, or the emergence of new platform features. Our analysis suggests that a single intervention rarely suffices; instead, multi-layered architectures that combine algorithmic diversity, dynamic fact-checking, and adaptive detection outperform any isolated strategy. However, layering multiple interventions introduces complex interactions. For example, increasing algorithmic diversity while simultaneously tightening fact-checking thresholds can create conflicting signals: users may receive a broader array of content but find much of it tagged as low credibility, leading to confusion and reduced trust in the platform as a whole [17]. Robust design requires careful calibration and empirical pilot testing before large-scale deployment.

Fairness concerns arise from the distributional effects of governance policies. Our simulations show that diversity-inducing algorithms disproportionately affect users with low tolerance for opposing viewpoints, causing some to leave the platform altogether. This exit may be a voluntary response to discomfort, but it can also be driven by targeted harassment if exposure to out-group content makes users more vulnerable to hostile interactions. Furthermore, bot-detection systems have historically shown bias against accounts from marginalized communities due to differences in communication style, network structure, and language use [18]. To mitigate such disparities, fairness-aware machine learning techniques should be incorporated into the detection pipeline, alongside oversight mechanisms that allow affected users to appeal automated decisions. Policymakers are increasingly calling for algorithmic auditing and external oversight boards to ensure that platform governance aligns with democratic values [19].

The policy implications extend beyond the platform level to the broader information environment. Government regulation, such as the European Union's Digital Services Act, imposes obligations on platforms to assess and mitigate systemic risks of misinformation. Our computational framework can serve as a testbed for evaluating specific regulatory proposals before implementation. For instance, a requirement to provide users with tools to control their own curation parameters (e.g., adjusting the diversity slider) may improve perceived control but could also reduce the effectiveness of system-level diversity enforcement if many users choose high personalization [20]. Similarly, transparency mandates that require platforms to disclose how algorithmic rankings are generated may enable adversarial gaming, as malicious actors can reverse-engineer the curation logic to amplify their content. These trade-offs need to be anticipated through rigorous modeling.

Finally, we emphasize that the computational approach does not replace qualitative social science or ethical deliberation but complements them. The quantitative insights derived from simulations must be interpreted within the context of political, cultural, and legal particularities. For example, a diversity-inducing policy that works well in a two-party system may behave differently in a multiparty democracy with strong regional media markets. Moreover, the very definition of misinformation is contested across jurisdictions; what constitutes harmful falsehood in one society may be considered protected speech in another. Therefore, the governance framework we advocate is not a one-size-fits-all technical fix but a process of iterative co-design among researchers, platform engineers, civil society groups, and regulators. The computational model supports this process by making explicit the assumptions and empirical consequences of different choices.

6. Conclusion

This paper has presented a computational study of misinformation diffusion and public opinion formation in online news networks, focusing on system-level architectural trade-offs and their implications for governance. Through a suite of agent-based simulations, we demonstrated that homophilic network structures and engagement-maximizing recommender systems amplify the spread of low-veracity content and deepen polarization. We evaluated several intervention strategies, including algorithmic diversity injection, proactive fact-checking, and adaptive bot detection, and identified both their benefits and their unintended consequences. The results underscore the importance of multi-layered, adaptable governance architectures that balance the competing goals of epistemic accuracy, user satisfaction, and fairness across demographic groups. Sustainability of these interventions requires ongoing investment in computational infrastructure and continuous learning to counter evolving adversary tactics.

Our work has several limitations that point to future research directions. The model abstracts away from the linguistic and multimodal nature of online content, treating information as label-annotated items; incorporating natural language processing into the simulation could capture more realistic cues for user beliefs and sharing decisions. Additionally, the current framework assumes a single platform, whereas real-world misinformation spreads across interconnected platforms with varying policies. Inter-platform modeling is a promising avenue for understanding cross-platform cascades and regulatory arbitrage. Finally, the calibration of behavioral parameters relies on existing empirical studies, but these studies often draw from Western, educated, industrialized, rich, and democratic populations.

Expanding the empirical base to diverse cultural contexts would enhance the generalizability of the findings. Despite these limitations, the computational approach offers a powerful lens for anticipating the systemic consequences of design choices and informing evidence-based policy in the critical domain of online news and democracy.

References

1. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.

2. Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39–50.

3. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096.

4. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113(3), 554–559.

5. Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., ... & Quattrociocchi, W. (2017). Debunking in a world of tribes. PLOS ONE, 12(7), e0181821.

6. Centola, D. (2010). The spread of behavior in an online social network experiment. Science, 329(5996), 1194–1197.

7. Barabási, A.-L. (2005). Network theory—the emergence of the creative enterprise. Science, 308(5722), 639–641.

8. Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. Journal of Consumer Research, 34(4), 441–458.

9. Starnini, M., Frasca, M., & Baronchelli, A. (2021). A high-sensitivity and low-cost method for the detection of coordinated inauthentic behavior in online social networks. Nature Communications, 12, 2598.

10. Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. Nature Communications, 10, 7.

11. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. Science, 363(6425), 374–378.

12. Solanki, D., Hsu, H. M., Zhao, O., Zhang, R., Bi, W., & Kannan, R. (2020, July). The way we think about ourselves. In International Conference on Human-Computer Interaction (pp. 276-285). Cham: Springer International Publishing.

13. Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1), 298–320.

14. Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132.

15. Schmidt, A., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2017). Anatomy of news consumption on Facebook. Proceedings of the National Academy of Sciences, 114(12), 3034–3039.

16. Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. Nature Communications, 9, 4787.

17. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.

18. Guess, A. M., Nyhan, B., & Reifler, J. (2019). Exposure to untrustworthy websites in the 2016 U.S. election. Nature Human Behaviour, 4, 472–480.

19. Ver Steeg, G., & Galstyan, A. (2012). Information transfer in social media. In Proceedings of the 21st International Conference on World Wide Web (pp. 509–518). ACM.

20. Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of Modern Physics, 81(2), 591–646.

21. Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q., & Vespignani, A. (2014). The Twitter of Babel: Mapping world languages through microbiogging platforms. PLOS ONE, 9(4), e93430.