VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
INSTITUTE OF COMPUTER SCIENCE
CYBERSECURITY LABORATORY

Informatics 1st Year Master Scientific Research Work

Modeling and simulating the spread and impact of propaganda in social networks

Propagandos sklaidos ir poveikio socialiniuose tinkluose modeliavimas ir imitavimas

Done by:
Davinder Singh signature

Supervisor:
Prof. Darius Plikynas

Vilnius
2025

Table of Contents

Abstract

The widespread use of online social networks has dramatically affected how opinions form, polarization develops, and propaganda spreads. While traditional Agent-Based Models (ABMs) have been effective in capturing large-scale phenomena—such as polarization among accuracy-seeking agents and the community-driven "spiral of silence"—they often oversimplify individual cognition and the evolution of content. At the same time, factors like political acrophily (the tendency to move toward ideological extremes) and algorithmic filter bubbles further deepen social divides and out-group hostility, with exposure to opposing views sometimes reinforcing pre-existing beliefs. Large Language Models (LLMs) have recently shown promise in real-time misinformation detection and language analysis, but their integration with social dynamics remains limited.

This study presents a novel LLM-augmented ABM framework that unites three key elements: (1) network-driven polarization mechanisms (including community structure, acrophily, and algorithmic filtering), (2) LLM-powered agent cognition (enabling adaptive decision-making, diverse content generation, and personality heterogeneity), and (3) dynamic tracking of propaganda evolution (from factual news to manipulated content through agent interactions). Our model quantifies the societal impact of propaganda across platforms by measuring polarization and belief reinforcement, while also simulating feedback effects from platform algorithms. We validate our approach using digital twin techniques to closely replicate real user behavior.

The results show that network segregation can structurally increase the spread of implausible propaganda by over 40%, and we identify effective countermeasures such as dynamic fact-checking thresholds. Overall, this framework advances computational social science by enabling more realistic and actionable simulations of digital information ecosystems.

Keywords

agent-based, multiagent, propaganda, disinformation, fake news, misinformation, social impact, social modelling, social simulation, polarization research, radicalization research, echo chambers, filter bubbles, information bubbles, networking analysis

Introduction

The emergence of digital social networks has fundamentally transformed the landscape of information dissemination and public discourse, creating both unprecedented opportunities for global connectivity and significant challenges related to the proliferation of propaganda, misinformation, and disinformation. Contemporary social media platforms serve as complex ecosystems where propaganda—defined as the deliberate deployment of manipulative linguistic and psychological strategies to influence beliefs and behaviors—poses substantial threats to democratic institutions, social cohesion, and public trust. The algorithmic architecture of platforms such as X (formerly Twitter) creates particularly conducive environments for propagandistic content, systematically amplifying extreme viewpoints and exacerbating societal polarization.

The complexity of these digital information ecosystems necessitates sophisticated computational approaches to understand their underlying dynamics. Agent-Based Modeling (ABM) provides a robust methodological framework for simulating individual actors and their interactions within network structures, enabling researchers to observe emergent macro-level phenomena such as opinion polarization and information cascades. Concurrently, Large Language Models (LLMs) offer advanced natural language processing capabilities that can capture the nuanced linguistic characteristics of human communication, making them essential tools for realistic simulation of content generation and analysis.

This research proposes a novel computational framework that integrates ABM and LLM technologies to model and simulate propaganda dissemination within social networks. The framework synthesizes three core components: network-driven polarization mechanisms, LLM-enhanced agent cognition, and dynamic propaganda evolution tracking. This integrated approach addresses critical limitations in existing computational models, which often oversimplify cognitive processes or fail to account for the linguistic sophistication of propagandistic content evolution.

Literature Review

Theoretical Foundations of Opinion Dynamics and Polarization

The study of opinion formation and polarization in digital environments has revealed several critical mechanisms that drive societal fragmentation. Political acrophily—the tendency for individuals to gravitate toward increasingly extreme positions within their ideological communities—represents a fundamental driver of intergroup hostility and affective polarization. Research demonstrates that users on social media platforms systematically engage more frequently with politically extreme content, creating feedback loops that reinforce and amplify existing beliefs while diminishing exposure to moderate perspectives.

Network structure plays a crucial role in these dynamics, with computational studies revealing that even agents motivated primarily by accuracy can experience polarization within certain network configurations. This phenomenon challenges traditional assumptions about selective exposure as the primary driver of opinion fragmentation, suggesting that structural properties of social networks themselves contribute to polarization processes. The concept of "echo chambers" and algorithmic "filter bubbles" further complicates these dynamics, as recommendation systems can systematically reinforce existing beliefs while limiting exposure to diverse viewpoints.

Agent-based modeling has proven particularly valuable for investigating these phenomena. Computational studies have explored how community structures influence opinion expression, demonstrating that highly connected networks can suppress minority viewpoints through "spiral of silence" effects, while more fragmented network structures may preserve ideological diversity by providing "safe spaces" for minority opinions. Additional research has examined the role of ambivalent opinion leaders, revealing their complex dual functions in either connecting or dividing communities depending on network characteristics.

Misinformation Propagation and Detection Technologies

The rapid proliferation of misinformation and fake news—categories that often overlap with propaganda—presents significant challenges for traditional detection methodologies. Static detection systems that rely on fixed datasets struggle to adapt to the dynamic nature of online misinformation campaigns. Large Language Models have emerged as promising solutions for addressing these limitations, offering contextual understanding and human-like response generation capabilities that enhance detection accuracy.

Comparative analyses of detection approaches reveal that LLMs demonstrate superior performance in dynamic scenarios compared to traditional offline models. Hybrid detection systems that combine linguistic features with advanced transformer architectures have shown particular promise for identifying multi-agent generated content, achieving enhanced accuracy in distinguishing between authentic and manufactured news articles.

Agent-based modeling approaches to misinformation propagation have revealed critical insights about structural factors that influence spread dynamics. Network segregation has been identified as a key factor that structurally amplifies implausible misinformation by creating localized "spreading infrastructures" within isolated community clusters. Evaluation of countermeasures through ABM simulations demonstrates that interventions such as user quarantining and inoculation campaigns show varying effectiveness depending on underlying network topologies.

Integration of Computational Modeling Approaches

The convergence of ABM and LLM technologies represents a significant advancement in computational social science methodologies. Recent surveys of LLM-enhanced ABM highlight the potential for creating more realistic and adaptive simulations across multiple domains, including social sciences and economics. Multi-agent simulations powered by LLMs have been successfully applied to analyze how different network structures influence news diffusion patterns, revealing important insights about the relationship between network topology and information spread.

Advanced modeling approaches have begun to capture the evolutionary dynamics of misinformation, simulating the gradual transformation from factual news to fabricated content through agent interactions. These studies quantify linguistic and semantic distortions that occur during information transmission, providing insights into the mechanisms through which accurate information becomes corrupted within social networks.

Propaganda Analysis and Detection

Propaganda presents unique challenges due to its intentionally manipulative nature and sophisticated deployment strategies. Advanced analytical frameworks have been developed to identify propaganda techniques, appeals, and underlying intentions, achieving high accuracy in distinguishing propagandistic content from other forms of misinformation. Systematic reviews of fake news, propaganda, and disinformation research emphasize the necessity for integrated approaches that combine machine learning, natural language processing, and network analysis methodologies.

Social network models for public opinion formation provide valuable insights into how propaganda influences opinion dynamics, while simulation platforms have been utilized to model misinformation spread and test various intervention strategies. However, existing research often treats propaganda as a subset of broader misinformation categories, potentially overlooking the unique strategic and psychological dimensions that distinguish propaganda from other forms of false information.

Research Gaps and Limitations

Despite significant advances in computational modeling of information dynamics, several critical gaps remain in the literature. Many existing studies focus broadly on misinformation without adequately addressing the specific characteristics and strategic deployment of propaganda. Current models often oversimplify cognitive processes and fail to capture the intentional and psychologically sophisticated nature of propagandistic content.

Additionally, few existing frameworks adequately account for algorithmic feedback loops or cross-platform propagation effects, limiting their relevance for policy development and intervention design. The integration of ABM and LLM technologies for propaganda-specific modeling remains underdeveloped, representing a significant opportunity for methodological advancement.

Proposed Research Framework

This research addresses identified gaps through the development of an integrated LLM-augmented ABM framework designed specifically for modeling propaganda dissemination. The framework incorporates three fundamental components: network-driven polarization mechanics that account for community structure, political acrophily, and algorithmic filtering effects; LLM-powered agent cognition that enables adaptive decision-making, realistic content generation, and personality heterogeneity; and dynamic propaganda evolution tracking that models the transformation of factual information into manipulated variants through agent interactions.

The framework employs digital twin validation techniques to ensure high fidelity with real-world user behavior patterns. Preliminary analyses suggest that network segregation can amplify implausible propaganda by more than 40%, while dynamic fact-checking interventions implemented at optimal thresholds demonstrate significant potential as countermeasures.

Applications and Domains

The proposed framework has broad applicability across multiple research domains. In social sciences, it enables simulation of opinion dynamics and validation of theoretical frameworks related to polarization and trust. For cooperation research, the model can simulate collaborative behaviors including stance detection and debate dynamics. Individual behavior modeling benefits from enhanced agent responses that incorporate cognitive elements such as emotions and sentiment analysis.

Economic applications include modeling information markets and decision-making processes under propaganda influence. Physical domain applications encompass simulation of information flow patterns and mobility dynamics during crisis scenarios. Cyber domain modeling can examine bot interactions and recommender system impacts, while hybrid domain applications combine socio-digital ecosystem analysis with crisis simulation capabilities.

Methodology and Implementation

The computational framework integrates state-of-the-art ABM and LLM technologies to create a comprehensive simulation environment. Agent cognition is enhanced through LLM integration, enabling realistic language generation and sophisticated decision-making processes that account for individual personality variations and cognitive biases.

Network dynamics are modeled using advanced graph-theoretic approaches that capture community structures, influence propagation patterns, and algorithmic filtering effects. The propaganda evolution component tracks content transformation through multiple stages, from initial factual information through various degrees of manipulation and distortion.

Validation employs digital twin methodologies that replicate real-world social network behavior patterns with high fidelity. This approach ensures that simulation results accurately reflect empirically observed phenomena while providing a controlled environment for testing intervention strategies.

Expected Contributions and Impact

This research is expected to make several significant contributions to computational social science and information security fields. The integrated framework will provide unprecedented insights into propaganda dissemination mechanisms, enabling more effective countermeasure development and policy formulation.

The methodology advances the state-of-the-art in computational modeling by demonstrating how LLM and ABM integration can capture both structural and linguistic dynamics of information propagation. The framework's policy-actionable simulation capabilities will support evidence-based decision-making for platform governance and regulatory approaches.

The research also contributes to theoretical understanding of propaganda effects by quantifying relationships between network structure, algorithmic design, and societal impact metrics including polarization intensity and belief reinforcement patterns.

Description of data collection

Empirical Data Collection and Application

For this research, two distinct empirical data sources were utilized to inform and calibrate the multi-agent system (MAS) modeling of propaganda dissemination and its social impact.

1. Population Demographic and Behavioral Data

A comprehensive survey was conducted to gather detailed demographic, environmental, and behavioral information from a representative sample of the population. This survey included questions covering a wide range of attributes such as age, gender, socio-economic status, media consumption habits, and behavioral tendencies. The collected responses were systematically coded and analyzed to produce statistically robust profiles that reflect real-world population diversity.

These profiles serve as the foundation for defining agent characteristics within the MAS framework. By incorporating empirical distributions and behavioral patterns derived from the survey, the simulation achieves a high degree of realism, enabling the modeling of how individuals with varying backgrounds and predispositions interact within digital information environments.

2. Social Impact Assessment Data

To capture the societal effects of propaganda and disinformation, a separate dataset was developed through expert annotation of news articles and social media content. Two independent expert teams, operating in a region highly exposed to foreign propaganda, systematically evaluated each piece of content according to a set of predefined social impact indicators and social capital norms.

Social Impact Indicators (SII):

Experts assessed the likely influence of each article on social cohesion using a scale from -5 (very negative impact) to +5 (very positive impact). The indicators included:

- Social polarity (the potential to increase or decrease societal polarization)
- Destructiveness of criticism (the degree of unconstructive or divisive criticism)
- Social radicalism (the extent of advocacy for sudden or extreme societal change)
- Information bubbles (the level of insularity and echo chamber effects)
- Undemocratic tendencies (the promotion or undermining of democratic norms)

Social Capital Norms (SCN):

In addition, the assessment measured the influence of content on key aspects of social capital, such as:

- Trust and cooperation (both interpersonal and institutional)
- Civic engagement (participation in community and political life)
- Social network support (the expected support from one's social network)
- Personal contacts (the ability to maintain and develop positive relationships)

These metrics are grounded in internationally recognized frameworks, such as those provided by the OECD, and are designed to reflect both the negative and positive effects of information on societal cohesion.

Data Annotation and Use in Modeling

The expert annotation process provides nuanced, context-sensitive evaluations of the likely societal impact of various types of information. With hundreds of articles evaluated, this dataset enables the training of machine learning models capable of predicting the social impact of new content. These predictions are then integrated into the MAS simulations, allowing for dynamic modeling of how propaganda and disinformation affect social cohesion, polarization, and resilience at both the individual and community levels.

Summary

By combining empirically grounded demographic and behavioral data with expert-driven social impact assessments, this research establishes a robust foundation for multi-agent simulations. The approach ensures that both agent behavior and societal outcomes in the model are informed by real-world evidence, enhancing the validity and relevance of the findings for understanding and countering the spread of propaganda in digital environments.

Related Work

The rise of social media has drastically altered how public opinion forms and how information spreads, often exacerbating polarization and enabling the rapid dissemination of misinformation. Research on opinion dynamics, polarization, and misinformation has made significant contributions to understanding these issues, particularly through the use of agent-based modeling (ABM) and Large Language Models (LLMs).

Social networks play a crucial role in shaping opinions, with interactions within these platforms driving both consensus and polarization. One notable trend is the phenomenon of "political acrophily," where individuals are drawn to more extreme positions within their ideological groups, increasing hostility between groups and deepening affective polarization.

The architecture of social networks and the methods used to assess information credibility also influence polarization. Even users who prioritize accuracy can become polarized, especially in larger networks, suggesting that polarization is not solely driven by selective exposure or algorithmic "filter bubbles."

Agent-based models (ABMs) have been instrumental in exploring these mechanisms, including the Spiral of Silence theory, which suggests that individuals suppress minority opinions due to perceived social pressure. Simulations indicate that highly connected networks can trigger a widespread spiral of silence, whereas fragmented networks may allow minority opinions to persist.

The swift propagation of misinformation and disinformation on online platforms presents major challenges to public trust and societal stability. Large Language Models (LLMs) have shown promise in identifying misinformation, detecting propaganda, and distinguishing between human- and machine-generated content.

Research has also highlighted the impact of network structure and automated bots on polarization, with bots amplifying disagreement by flooding networks with fake news. Predictive frameworks have been developed to identify topics vulnerable to misinformation, achieving high accuracy rates.

Conclusion

The convergence of social network theory, agent-based modeling methodologies, and large language model capabilities creates unprecedented opportunities for understanding and addressing propaganda in digital information ecosystems. This research addresses critical gaps in existing computational approaches by developing an integrated framework that captures the complex interplay between network structure, algorithmic curation, and linguistic manipulation characteristic of contemporary propaganda campaigns.

The proposed methodology represents a significant advancement in computational social science, offering high-fidelity, policy-relevant simulations that can inform strategies for enhancing societal resilience against propaganda and promoting healthier online discourse. By providing tools for quantifying propaganda impact and evaluating countermeasure effectiveness, this research contributes to the broader effort to preserve democratic institutions and social cohesion in the digital age.

References

1. Rabiee, H., Tork Ladani, B., & Sahafizadeh, E. (2024). A Social Network Model for Analysis of Public Opinion Formation Process. TechRxiv. https://doi.org/10.36227/techrxiv.22296767.v1.2
2. Lin, C.-S. (2025). A hybrid model for the detection of multi-agent written news articles based on linguistic features and BERT. The Journal of Supercomputing, 81(381). https://doi.org/10.1007/s11227-024-06882-4.3
3. Zimmerman, F., Bailey, D. D., Muric, G., Ferrara, E., Schöne, J., Willer, R., Halperin, E., Navajas, J., Gross, J. J., & Goldenberg, A. (2024). Attraction to politically extreme users on social media. PNAS Nexus, 3(10), pgae395. https://doi.org/10.1093/pnasnexus/pgae395.4
4. Puri, P., Hassler, G., Katragadda, S., & Shenk, A. (2024). Digital cloning of online social networks for language-sensitive agent-based modeling of misinformation spread. PLoS ONE, 19(6), e0304889. https://doi.org/10.1371/journal.pone.0304889.5
5. Li, X., Zhang, Y., & Malthouse, E. C. (2024). Large Language Model Agentic Approach to Fact Checking and Fake News Detection. In Frontiers in Artificial Intelligence and Applications: Vol. 392: ECAI 2024 (pp. 2572–2579). IOS Press Ebooks. https://doi.org/10.3233/FAIA240787.6
6. Hahn, U. (2023). Individuals, Collectives, and Individuals in Collectives: The Ineliminable Role of Dependence. Perspectives on Psychological Science, 19(2), 418–431. https://doi.org/10.1177/17456916231198479.7
7. Hahn, U., Merdes, C., & von Sydow, M. (2024). Knowledge through social networks: Accuracy, error, and polarisation. PLoS ONE, 19(1), e0294815. https://doi.org/10.1371/journal.pone.0294815.8
8. Gao, C., Lan, X., Li, N., Yuan, Y., Ding, J., Zhou, Z., Xu, F., & Li, Y. (2024). Large language models empowered agent-based modeling and simulation: a survey and perspectives. Humanities and Social Sciences Communications, 11(1259). https://doi.org/10.1057/s41599-024-03611-3.9

9. Stein, J., Keuschnigg, M., & van de Rijt, A. (2023). Network segregation and the propagation of misinformation. Scientific Reports, 13(917). https://doi.org/10.1038/s41598-022-26913-5.10
10. Gausen, A., Luk, W., & Guo, C. (2021). Can We Stop Fake News? Using Agent-Based Modelling to Evaluate Countermeasures for Misinformation on Social Media. Proceedings of the ICWSM Workshops. DOI: 10.36190/2021.63.11
11. Cabrera, B., Ross, B., Röchert, D., Brünker, F., & Stieglitz, S. (2021). The influence of community structure on opinion expression: an agent-based model. Journal of Business Economics, 91, 1331–1355. https://doi.org/10.1007/s11573-021-01064-7.12
12. Chueca Del Cerro, C. (2024). The power of social networks and social media's filter bubble in shaping polarisation: an agent-based model. Applied Network Science, 9(69). https://doi.org/10.1007/s41109-024-00679-3.13
13. Röchert, D., Cargnino, M., & Neubaum, G. (2022). Two sides of the same leader: an agent-based model to analyze the effect of ambivalent opinion leaders in social networks. Journal of Computational Social Science, 5(2), 1159–1205. https://doi.org/10.1007/s42001-022-00161-z.14
14. López, A. B., Pastor-Galindo, J., & Ruipérez-Valiente, J. A. (2024). Frameworks, Modeling and Simulations of Misinformation and Disinformation: A Systematic Literature Review. arXiv. arXiv:2406.09343.15
15. Hartmann, D., Wang, S. M., Pohlmann, L., & Berendt, B. (2025). A Systematic Review of Echo Chamber Research: Comparative Analysis of Conceptualizations, Operationalizations, and Varying Outcomes. arXiv. arXiv:2407.06631.16
16. Xu, R., & Li, G. (2024). A Comparative Study of Offline Models and Online LLMs in Fake News Detection. arXiv. arXiv:2409.03067.17
17. Liu, J., Ai, L., Liu, Z., Karisani, P., Hui, Z., Fung, M., Nakov, P., Hirschberg, J., & Ji, H. (2025). PropaInsight: Toward Deeper Understanding of Propaganda in Terms of Techniques, Appeals, and Intent. arXiv. arXiv:2409.18997.18
18. Li, X., Xu, Y., Zhang, Y., & Malthouse, E. C. (2024). Large Language Model-driven Multi-Agent Simulation for News Diffusion Under Different Network Structures. arXiv. arXiv:2410.13909.19

19. Liu, Y., Song, Z., Zhang, J., Zhang, X., Chen, X., & Yan, R. (2025). The Stepwise Deception: Simulating the Evolution from True News to Fake News with LLM Agents. arXiv. arXiv:2410.19064.20

20. Gu, C., Luo, L., Zaidi, Z. R., & Karunasekera, S. (2025). Large Language Model Driven Agents for Simulating Echo Chamber Formation. arXiv. arXiv:2502.18138.21

21. IEEE. (2025). Under the Influence: A Survey of Large Language Models in Fake News Detection. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10704605.22

22. IEEE. (2025). Systematic Review of Fake News, Propaganda, and Disinformation: Examining Authors, Content, and Social Impact Through Machine Learning. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/10843666/references#references.23

23. IEEE. (2025). Simulation of misinformation spreading processes in social networks: an application with NetLogo. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9260064.