23rd International Conference on Modelling and Applied Simulation, 022
21st International Multidisciplinary Modeling & Simulation Multiconference

2724-0061 © 2024 The Authors.
doi: 10.46354/i3m.2024.mas.022

Innovative Modeling of Deep Fake Dynamics on the Population by using Artificial Neural Network

Agostino G. Bruzzone[1,2,3*], Antonio Giovannetti[2,3], Luca Cirillo[3], Filippo Ghisi[2,3], Simone Gleisson Ferrero[2,3]

[1]DIME, Genoa University, Via Opera Pia 15, Genova, 16145, Italy
[2]SIM4Future, Via Trento 43, 16145 Genova, Italy
[3]Simulation Team, Via Cadorna 2, Savona, 17100, Italy

[*]Corresponding author. Email address:agostino.bruzzone@simulationteam.com

Abstract

This paper presents an innovative model for understanding the dynamics of information propagation within social networks. Incorporating cognitive biases, follower influence, and temporal decay, we propose a mathematical framework to simulate how information spreads through a network of individuals connected by varying degrees of trust, familiarity, and social influence modeled as a Neural Network. Our model accounts for the role of confirmation bias, the bandwagon effect, and fact-checking delays to capture real-world phenomena that affect the spread of true and false information alike. This innovative model is based on a hybrid approach that uses components based both on static Neural Graphs (to capture the structure of the social network) and on models of epidemic diffusion of information (to model the dynamics of propagation over time). To test the model we used LLMs and open source data to generate opinions respect to different topics in the population network based on different factors, such as age, gender, social status, educational level etc. The authors introduce in the network different messages (real and fake) trough an embedding layer in order to understand the spread of information. The proposed model is validated against state-of-the-art approaches and aims to enhance predictive accuracy in fields such as misinformation control and viral marketing.

Keywords: Modeling & Simulation; Human Behavior Modeling; Large Language Models; Social Media Analysis

1. Introduction

The propagation of information in social networks is a critical field of study due to its widespread implications in areas such as politics, marketing, and public health. In an interconnected world, understanding how information, whether factual or false, scatters across populations are crucial to mitigate the spread of misinformation and optimize information campaigns. This paper aims to propose an innovative model for simulating information propagation by integrating cognitive and social factors, including biases, trust levels, and the influence of followers.

In recent years, a wide variety of models have been developed to study how information spreads across social networks, but few effectively capture the complexity of human decision-making and social interactions. Our proposed model aims to improve upon current methods by introducing new elements such as cognitive bias-driven behavior and dynamic updates to trust levels based on the credibility of the information received.

Traditional models of information propagation, such as Epidemic Diffusion Models (e.g., SIR—Susceptible,

Infected, Recovered), have been widely used in the study of information the spread, however, these models are typically limited in their ability to account for the nuanced behavioral and cognitive factors that affect how individuals receive, interpret, and propagate information. Cognitive biases, such as confirmation bias and the bandwagon effect, play a critical role in shaping an individual's likelihood of believing and spreading certain types of information. Additionally, the level of trust in the source, the delay in fact-checking, and the influence of a user's social network (e.g., followers) further complicate the dynamics of information flow.

To address these challenges, we propose a novel hybrid model that integrates Graph Neural Networks (GNNs) with an extended SIR diffusion model, incorporating cognitive biases and social influence factors. Unlike traditional unidimensional models of information flow, our approach captures the multidimensional nature of information propagation, where the beliefs, trustworthiness, and fact-checking capabilities of individuals influence not only whether they propagate information but also how quickly they verify its truthfulness. Our model represents the social network as a directed graph, where each node corresponds to an individual, and edges represent the connections between them. Each node is characterized by a set of attributes, including, among others, educational level, cognitive biases, trustworthiness, fact-checking delay, and the number of followers. The interactions of these attributes determine whether a node propagates received information based on an activation function that is sensitive to confirmation bias, social influence, and the perceived truthfulness of the information.

Furthermore, the model leverages a Dynamic Graph Neural Network (DGN) to account for evolving social connections and changing states over time. The use of Graph Attention Networks (GAT) enables the model to dynamically adjust the influence of neighboring nodes based on their attributes, further refining the propagation process. Additionally, the SIR states (Susceptible, Infected, Recovered) of the nodes evolve in response to the aggregated information received from neighboring nodes, which is moderated by the cognitive and social factors.

2. State of the art

The study of information propagation in social networks has evolved significantly over the last decade. Traditional models, such as the Independent Cascade (IC) model and the Linear Threshold (LT) model, focus primarily on network structure and the probabilities of nodes activating based on the state of their neighbors, often overlooking the psychological and social factors that influence individual decision-making. In fact, these models present some limitations, systematically underestimating the spreading speed and randomness of information (Ran et al., 2020), or are based on strong assumption, such as that the transmission of information is not affected by the behavior of other users (He et al., 2024), that hinder their ability to reliably represent real social diffusion dynamics.

The integration of intelligent agents and human behavior models plays a pivotal role in simulating information spreading, particularly in complex and dynamic scenarios, such as urban riots or large-scale social events. Intelligent agents, as highlighted by Bruzzone et al. (2014a, 2014b), are capable of autonomously simulating the decision-making processes and interactions of individuals within these environments. These agents are designed to mimic human behavior by considering cognitive, emotional, and social factors, which are essential for understanding how information propagates through a population.

The use of such agents enables a more accurate representation of the emergent behaviors that arise in crisis situations, where the rapid dissemination of information—whether through formal channels or social networks—influences public perception and collective actions. Furthermore, the work of Bruzzone et al. (2011) emphasizes the effectiveness of these agents in driving computer-generated forces to simulate human behavior in urban riots, showcasing their applicability in modeling large-scale social phenomena. By simulating human behavior through intelligent agents, researchers gain deeper insights into how information is spread in both controlled and chaotic settings, thereby enhancing decision support systems for emergency response and crowd management. Some recent efforts have attempted to integrate cognitive biases into propagation models (Neuhäuser et al., 2021; Ecker et al., 2022), accounting for how individuals' preexisting beliefs affect their likelihood of adopting new information. Confirmation bias in particular (Mao, Akyol and Hovakimyan, 2021), plays a pivotal role, as individuals tend to favor information that aligns with their existing beliefs, often dismissing contradictory evidence. Similarly, the bandwagon effect describes how the likelihood of adopting information increases as more individuals within a network embrace it.

Other advancements have focused on the role of social influence (Li, Zhang, Huang, 2018), where the number of followers or the trustworthiness of connections impacts how quickly information spreads. For instance, models that account for trust between individuals provide a more nuanced view of information flow, as people are less likely to adopt information from sources they deem untrustworthy. Although these efforts mark significant progress, most models fail to incorporate all necessary elements (Chen, Xiao, Kumar, 2023), such as cognitive biases, the decay of information over time, and the dynamic nature of social connections. This paper seeks to address these gaps by integrating multiple factors into a unified model.

3. Conceptual Model

Our model represents a social network as a directed graph G=(V,E), where nodes V represent individuals and edges E represent the connections between them. Each node possesses various attributes that influence its decision to propagate information, including:

- Educational level (e_i)
- Cognitive biases (b_i)
- Trustworthiness (t_{ri})
- Fact-checking delay (r_i)
- Number of followers (f_i)
- Beliefs (B_i)
- Information truthfulness (v_i)

These attributes interact to determine the activation function of node i at time t, which dictates whether the node propagates the received information to its neighbors.

Each node also represents an individual inside a group of people, thus it is also described by several parameters that characterize a person, like:

- Age
- Gender
- Social Status
- Consensus
- Political Orientation
- Education Level
- Religion

A message (information) is sent into the network through an embedding process. Each node generates through an LLM an opinion based on the different characteristic of the person, and the embedding layer compute the values of the different opinions. Based on the parameters of each node that is reached by the information, a level of Consensus is calculated to determine how similar it is to the general Values and Beliefs of the subject, directly effecting the influence of the Cognitive Biases (people are more likely to believe and diffuse information that align with their personal views, especially on socio-economics and political themes) and the spread of the information from that node. In similar way it is computed the bandwagon effect. In the next session the authors describe how the information spread through the network and how the connections among nodes change.

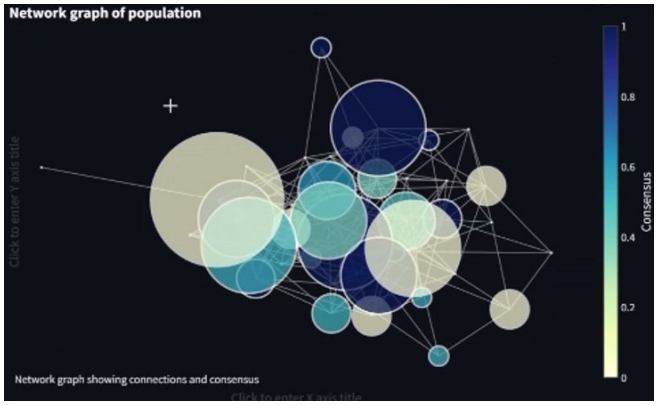

Figure 1. Network Graph and Population Consensus

4. General Architecture

The presented hybrid model integrates static graph neural networks (GNNs) with an epidemic diffusion model (SIR) and cognitive biases to simulate the propagation of information (e.g., fake news) through a social network. This model captures both the cognitive and social influences on information spread, with a dynamic update of node states and connections over time.

Social Network Creation (Graph Representation)

Input: A social network is represented as a graph G=(V,E) where:

- V is the set of nodes, each representing an individual
- E is the set of edges, representing social connections among individuals

Each node i is associated with a vector of features that describe the behavior of the node:

- Cognitive biases (b_i): Inclination to believe information that aligns with personal convictions.
- Trustiness (t_{ri}): The level of trust a node has in the information it receives.
- Fact-checking delay (r_i): The time it takes for a node to verify information, which is directly related to the educational level of the person
- Personal beliefs (B_i): The pre-existing beliefs of the node.
- SIR state: Each node begins in the S (Susceptible) state, while some nodes may be initialized in the I (Infected) state

SIR Model Integration

- The nodes of the network are in one of the following three states:
  - Susceptible (S): A node that has not yet received the information.
  - Infected (I): A node that has received the information and is propagating it.

    - Recovered (R): A node that has verified the information, stopped propagating it, or corrected it.
- Transitions between SIR states are influenced by:
    - Cognitive biases: A node with a strong confirmation bias or bandwagon effect is more likely to become infected when receiving information that aligns with its beliefs.
    - Trustiness and fact-checking: Nodes with high trustiness and slower fact-checking are more likely to transition to the recovered state if they verify the information as false.

Dynamic Graph Neural Network (DGN)

- The model incorporates a Dynamic Graph Neural Network (DGN) to update the representations of nodes and their social connections over time.
- Graph Attention Networks (GAT) are used to dynamically calculate the weight and influence of neighboring nodes. Each node has a representation $h_{i,t}$, which evolves over time based on the information it receives and its own social and cognitive parameters.
- At each time step t, the representation of each node is updated through an aggregation function that combines:
    - Information received from neighboring nodes.
    - Local node features (cognitive biases, trustiness, etc.).

SIR Transitions in the Neural Network

- Each node's SIR state evolves over time based on aggregated information from its neighbors.
- Transitions between states (from Susceptible to Infected or from Infected to Recovered) are modulated by cognitive biases and information verification:
    - A node in the Infected state may transition to the Recovered state if it verifies that the information is false.
    - A node in the Susceptible state could become Infected if neighboring nodes propagate information that aligns with its cognitive biases.

Information Propagation

- Information propagation within the network is regulated by an activation function that determines whether a node propagates the received information:
    - The probability of propagation is influenced by confirmation bias and the bandwagon effect (propagating information because many neighbors do so).
- Information is transmitted from node to node, and the weights of the social connections are dynamically updated, with the weight decreasing over time through a temporal decay factor.

Network Update and Temporal Simulation

- The simulation runs over multiple iterations. In each iteration:
    - The states of the nodes (S, I, R) are updated.
    - The weights of the connections between nodes are modified based on the accuracy and quality of the information transmitted.
    - Cognitive biases and social influences modulate the information propagation process.
- At the end of the simulation, we obtain the overall diffusion of the information (i.e., how many nodes were infected and how many recovered).

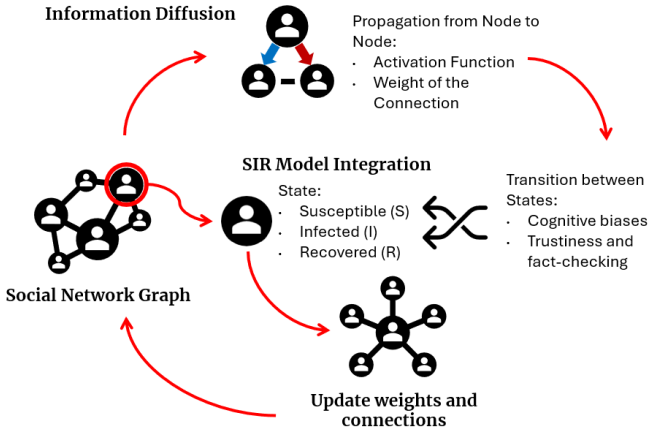

Figure 2. Dynamic Graph Neural Network General Architecture

5. Mathematical Formulation

Each node $i \in V$ is associated with a feature vector x_i, including:

$$x_i = (b_i, t_{ri}, r_i, B_i, S_i, I_i, R_i)$$

Where:

- b_i: Cognitive biases.
- t_{ri}: Trustiness.
- r_i: Fact-checking delay.
- B_i: Personal beliefs.
- S_i, I_i, R_i: SIR state (Susceptible, Infected, Recovered).

The update of the state of each node i at time t is calculated through an aggregation function:

$$h_{i,t+1} = \sigma(W \cdot \sum_{j \in N_i} a_{ij,t} h_{j,t} + W' \cdot h_{i,t})$$

Where:

- $N(i)$ is the set of neighbors of node i.
- $a_{ij,t}$ is the attention coefficient between nodes i and j, which weights the influence of neighbors.
- $h_{i,t}$ is the representation of node i at time t.
- W and W' are weight matrices learned by the network.

SIR Transitions in the GNN

The transition of state for node i at time t+1 depends on the aggregated state of the neighbors:

$$h_{i,t+1} = \sigma(W \cdot \sum_{j \in N_i} a_{ij,t} h_{j,t} + W' \cdot h_{i,t}) \cdot (1 - \gamma_{i,t}) + R_{i,t} \cdot \gamma_{i,t}$$

Where:

- $\gamma_{i,t}$ is the recovery rate, representing the probability that an infected node transitions to the recovered state.

Activation Function for Information Propagation

The probability that node iii propagates information I at time t is governed by an activation function:

$$f(I_i, t) = \sigma(w_{ij}^{tipo} \cdot I_i \cdot tr_i \cdot v_i - \theta_i) \cdot (1 + \alpha \cdot sim(M, B_i)) \cdot (1 + \beta \cdot \left(\frac{n_{propaganti}}{n_{totali}}\right)) \cdot e^{-\mu t}$$

Where:
σ is the sigmoid function .
- $w_{ij} = w_{ij}^{tipo} \cdot tr_i$ is the weight associated with node I and node j (the connection could be family, social, friend or acquaintance).
- I_i is the information received from node i.
- θ_i is the activation threshold.
- α e β are parameters of intensity of confirmation bias and bandwagon effect.
- $sim(M, B_i)$ is the similarity between message M and beliefs B_i of node i.
- n_propagants is the number of neighbors who have already propagated the message.
- n_totals is the total number of neighbors of the node.
- $e^{-\mu t}$ is the time decay term with parameterμ.

Social $= w_{ij}^{social} \cdot (1 + \gamma \cdot f_i)$,γ regulates influence of followers

The similarity between the message M and the beliefs B_i of the node is computed as:

$$sim(M, B_i) = \frac{M \cdot B_i}{||M|| ||B_i||}$$

- M and B_i are the vectors of message and belief characteristics, respectively.
- ||M|| and $||B_i||$ are the carriers' rules.

If a node receives false information, it increases the strength of the connection towards the node that gave the false information by reducing the weight of that connection in the future:

$$w_{ij,nuovo} = w_{ij} \cdot \delta$$

where δ is a reduction factor (0<δ<1).

6. Results and Discussion

The preliminary tests of our hybrid model, which integrates graph neural networks (GNNs) with the SIR diffusion model to simulate the propagation of information through a social network, were assessed using several key metrics, including the rate of information spread, the influence of cognitive biases, and the effectiveness of fact-checking in limiting the spread of misinformation. Furthermore, we validated the model through a combination of synthetic experiments and real-world comparison with known information dissemination patterns. We simulated the model on a synthetic social network generated using an Erdős–Rényi random graph. The network contains 1,000 nodes, where each node represents an individual and edges represent social connections. Each node is initialized with attributes such as cognitive biases, trustworthiness, fact-checking delay, and the number of followers. The connections between the nodes are defined as either strong or weak, with different initial edge weights representing varying levels of social influence.

The simulation was conducted over 100 epochs, where information is introduced into the network via a randomly selected seed node. This node starts in the infected state and propagates the information according to the activation function, influenced by confirmation bias, bandwagon effect, and temporal decay. We conducted multiple experiments by varying the parameters of the model, such as the bias coefficient α, social influence β, and recovery rate γ, to assess their impact on the overall diffusion process.

We observed that the rate of infection increases rapidly during the early epochs as the information spreads through nodes with high cognitive biases and strong social connections. The confirmation bias (α) has a significant impact on the speed and extent of propagation, with higher values leading to faster and more widespread infection.

As expected, nodes with lower trustworthiness and longer fact-checking delays remained infected for longer periods, contributing to sustained

misinformation spread. In contrast, nodes with higher trust and more rigorous fact-checking transitioned to the recovered state more quickly, helping to mitigate the spread of false information. The inclusion of temporal decay (μ) further slowed the spread of misinformation in later epochs, as information becomes outdated and loses its influence.

The experiments demonstrate the relationship between the fact-checking delay (r_i) and the proportion of nodes that transition to the recovered state. As expected, nodes with shorter fact-checking delays are more likely to verify the information and recover from the infected state.

Furthermore, we observed that the overall recovery rate (γ) increases when nodes have a high degree of social trust and are part of stronger social connections. This indicates that fact-checking is more effective in trusted networks, where individuals are more likely to rely on the information provided by their close social contacts.

7. Conclusions

The hybrid model presented in this paper integrates Graph Neural Networks (GNNs) with an extended SIR diffusion model, offering a novel approach to simulating the propagation of information, including misinformation, within social networks. Through the incorporation of key social and cognitive factors such as confirmation bias, trustworthiness, fact-checking delay, and social influence, this model captures the multidimensional nature of how individuals process and spread information and fake news. The findings of this research provide several avenues for further investigation and practical applications. From a research perspective, the model could be extended to incorporate more nuanced factors such as emotional engagement, political orientation, or geographical clustering. Additionally, future studies could apply the model to more specific real-world datasets, allowing for more precise validation and refinement of the model's parameters.

References

Bruzzone, A., Massei, M., Longo, F., Poggi, S., Agresta, M., Bartolucci, C. & Nicoletti, L. (2014a). Human behavior simulation for complex scenarios based on intelligent agents. In Proceedings of the 2014 Annual Simulation Symposium (pp. 1 10).

Bruzzone, A., Massei, M., Longo, F., Poggi, S., Agresta, M., Bartolucci, C., & Nicoletti, L. (2014b, April). Human behavior simulation for complex scenarios based on intelligent agents. In *Proceedings of the 2014 Annual Simulation Symposium* (pp. 1-10).

Bruzzone, A. G., Tremori, A., Tarone, F., & Madeo, F. (2011). Intelligent agents driving computer generated forces for simulating human behaviour in urban riots. *International Journal of Simulation and Process Modelling*, *6*(4), 308-316.

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment analysis in social networks*. Morgan Kaufmann.

Qi, J. (2024, January). The Impact of Large Language Models on Social Media Communication. In *Proceedings of the 2024 7th International Conference on Software Engineering and Information Management* (pp. 165-170).

Pew Research Center, "News Consumption Across Social Media in 2021"

GlobalWebIndex, "The Global Media Landscape in 2023"

Aarøe, L., & Petersen, M. B. (2020). Cognitive biases and communication strength in social networks: The case of episodic frames. *British Journal of Political Science*, *50*(4), 1561-1581.

Ran, Y., Deng, X., Wang, X., & Jia, T. (2020). A generalized linear threshold model for an improved description of the spreading dynamics. *Chaos, 30 8*, 083127.

He, Jixuan & Guo, Yutong & Zhao, Jiacheng. (2024). Exploring the Independent Cascade Model and Its Evolution in Social Network Information Diffusion.

Neuhäuser, L., Stamm, F.I., Lemmerich, F. *et al.* Simulating systematic bias in attributed social networks and its effect on rankings of minority nodes. *Appl Netw Sci* **6**, 86 (2021).

Sijing Chen, Lu Xiao, Akit Kumar (2023). Spread of misinformation on social media: What contributes to it and how to combat it. Computers in Human Behavior, Volume 141, 107643, ISSN 0747-5632.

Y. Mao, E. Akyol and N. Hovakimyan, "Impact of Confirmation Bias on Competitive Information Spread in Social Networks", in *IEEE Transactions on Control of Network Systems*, vol. 8, no. 2, pp. 816-827, June 2021, doi: 10.1109/TCNS.2021.3050117.

Ecker, U.K.H., Lewandowsky, S., Cook, J. *et al.* The psychological drivers of misinformation belief and its resistance to correction. *Nat Rev Psychol* **1**, 13–29 (2022).

Kan Li, Lin Zhang, Heyan Huang (2018). Social Influence Analysis: Models, Methods, and Evaluation, Engineering, Volume 4, Issue 1, Pages 40-46, ISSN 2095-8099.