JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 2 (2025), pp. 86–101
DOI: 10.63646/jaiaa.2025.030205

AI-Enhanced Prediction of Collective Opinion Shifts and Cooperative Behavior Diffusion in Online Social Networks

Wei Zhang[1], Lin Xu[2], Xiaomei Chen[3,*]

[1] School of Computer Science and Technology, Zhengzhou University of Light Industry, Zhengzhou 450002, China

[2] Department of Information Management, Shandong University of Science and Technology, Qingdao 266590, China

[3] School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin 541004, China

* Corresponding author: chenxm@guet.edu.cn

ARTICLE INFO
Received
January 18, 2024
Revised
March 21, 2024
Accepted
May 12, 2025
Available Online
June 30, 2025
DOI
10.63646/jaiaa.2025.030205

Publisher
INATGI, United States of America
Journal
JAIAA – ISSN 3067-7386

Abstract
The interplay between individual opinion formation and cooperative behavioral choices in online social networks represents a critical yet underexplored dimension of computational social science. This study presents an AI-enhanced dual-layer network framework that integrates mathematical opinion-behavior co-evolution modeling with state-of-the-art machine learning techniques to predict collective opinion shifts and cooperative behavior diffusion. The opinion layer adopts a weighted DeGroot-like update rule, while the behavior layer synthesizes three social influence mechanisms—neighbor imitation, payoff-driven decision-making, and cognitive consistency pressure. Building upon this foundation, we develop a graph convolutional network (GCN) architecture with graph attention (GAT) aggregation to capture topological opinion propagation patterns, a proximal policy optimization (PPO)-based reinforcement learning agent for cooperative strategy optimization, and an ensemble of machine learning classifiers for opinion shift detection. SHAP (SHapley Additive exPlanations) analysis is applied to quantify the contribution of each network and behavioral feature to model predictions. Experiments across scale-free, random, and small-world network topologies reveal that scale-free structures accelerate opinion convergence and sustain higher cooperation ratios under both synchronous and asynchronous update mechanisms. The GCN-GAT model achieves a mean absolute error (MAE) of 0.031 for opinion prediction, outperforming baseline LSTM and random forest models. Empirical validation against longitudinal survey data on environmental attitudes and pro-environmental behaviors among Chinese social media users demonstrates strong concordance between model trajectories and observed trends. This work advances the understanding of opinion-behavior co-evolution by providing an interpretable, data-driven prediction framework applicable to public opinion monitoring, social intervention design, and behavioral policy assessment.

Keywords: Social networks; Opinion dynamics; Cooperative behavior; Graph neural networks; Reinforcement learning; SHAP explainability; Dual-layer network; Co-evolution

I. INTRODUCTION

The proliferation of online social platforms has fundamentally transformed how individuals form, express, and revise opinions.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

Unlike traditional face-to-face interactions, digital environments enable rapid propagation of views across geographically dispersed populations, creating unprecedented opportunities for both collective intelligence and collective polarization. Understanding how group opinions evolve and how cooperative behaviors spread within these networks has become a central challenge in computational social science (Castellano et al., 2009; Watts and Strogatz, 1998). Recent empirical studies have demonstrated that online exposure to diverse opinions can shift individual attitudes and ultimately alter aggregate behavioral patterns, especially in domains such as environmental awareness, public health, and civic participation (Bakshy et al., 2015; Centola, 2010; Vosoughi et al., 2018).

Existing research on opinion dynamics has largely followed two parallel tracks. The first focuses on opinion evolution using mathematical update rules, exemplified by the seminal DeGroot model of weighted consensus formation (DeGroot, 1974), the bounded-confidence models of Deffuant et al. (2000) and Hegselmann and Krause (2002), and subsequent extensions including stubborn agent models (Ghaderi and Srikant, 2014) and social influence studies (Moussaid et al., 2013). The second track addresses behavioral decision-making through evolutionary game theory, examining how cooperative strategies emerge and diffuse in networked populations (Axelrod and Hamilton, 1981; Nowak and May, 1992; Santos and Pacheco, 2005; Perc et al., 2017). The critical limitation of these existing approaches is that they treat opinion and behavior as independent processes, despite ample empirical evidence that they co-evolve through complex feedback mechanisms (Macy and Willer, 2002; Flache et al., 2017).

The advent of artificial intelligence offers powerful new tools for addressing this fundamental limitation. Graph neural networks (GNNs) can learn topological representations of social networks directly from data, capturing structural regularities that govern information propagation (Kipf and Welling, 2017; Hamilton et al., 2017; Velickovic et al., 2018; Zhou et al., 2020). Reinforcement learning (RL) provides a framework for understanding and optimizing cooperative strategies in dynamic environments (Mnih et al., 2015; Schulman et al., 2017). Machine learning classification enables systematic identification of conditions under which opinion shifts occur, while SHAP (SHapley Additive exPlanations) analysis offers post-hoc interpretability by quantifying each feature's marginal contribution to model predictions (Lundberg and Lee, 2017; Ribeiro et al., 2016). Despite the clear potential of these AI methods, their systematic integration into dual-layer opinion-behavior co-evolution models remains largely unexplored.

Building upon the dual-layer network modeling paradigm (Boccaletti et al., 2014; Kivelä et al., 2014) and recent advances in social simulation (Quattrociocchi et al., 2014; Dong et al., 2017), this paper makes four core contributions. First, we formalize a dual-layer network model where the opinion layer adopts a DeGroot-like weighted update rule and the behavior layer integrates three social influence mechanisms: neighbor imitation, payoff-driven decision-making, and cognitive consistency pressure. Second, we develop an AI-enhanced prediction framework incorporating a GCN-GAT model for opinion trajectory prediction, a proximal policy optimization (PPO)-based RL agent for cooperative behavior optimization, and an ensemble of machine learning classifiers for opinion shift detection. Third, we apply SHAP analysis to decompose feature contributions and provide mechanistic interpretability. Fourth, we validate the integrated model against longitudinal survey data on environmental attitudes and pro-environmental behaviors, demonstrating that model trajectories align closely with real-world observed trends.

The remainder of this paper is organized as follows. Section II reviews related work on opinion dynamics, evolutionary game theory, and AI applications in social system analysis. Section III formalizes the dual-layer network model for opinion-behavior co-evolution. Section IV describes the AI-enhanced prediction framework. Section V presents SHAP-based interpretability analysis. Section VI reports simulation experimental results. Section VII validates the model against empirical survey data. Section VIII provides discussion, and Section IX concludes the paper.

II. RELATED WORK

A. Opinion Dynamics Models

The mathematical modeling of opinion dynamics has a rich history spanning sociology, physics, and computer science. DeGroot (1974) introduced the foundational weighted averaging model in which agents iteratively update opinions based on the weighted average of neighbors' beliefs, establishing conditions for asymptotic consensus. Friedkin and Johnsen (1990) extended this by incorporating stubborn agents with fixed internal opinions, producing richer steady-state distributions. The bounded-confidence framework independently proposed by Deffuant et al. (2000) and Hegselmann and Krause (2002) introduced the key mechanism that agents only influence each other when their opinions are within a given tolerance threshold, enabling the emergence of opinion clusters and polarization. Subsequent developments incorporated network topology effects: Sood and Redner (2005) analyzed the voter model on heterogeneous networks, while Flache et al. (2017) provided a comprehensive review identifying frontiers including

network feedback and multi-dimensional opinion spaces. Lorenz (2007) offered a systematic survey of continuous opinion models, and Acemoglu and Ozdaglar (2011) examined opinion learning equilibria in social networks through a game-theoretic lens.

B. Evolutionary Game Theory and Cooperation

The emergence and diffusion of cooperative behavior in social dilemmas has been extensively studied through evolutionary game theory. Axelrod and Hamilton (1981) demonstrated that cooperation can evolve through reciprocity even among self-interested agents. Nowak and May (1992) showed that spatial structure alone can sustain cooperation in prisoner's dilemma games without additional mechanisms. Santos and Pacheco (2005) established that scale-free network topology provides a unifying framework for the spontaneous emergence of cooperation, a result that has since been replicated and extended in numerous studies (Perc and Szolnoki, 2010; Szolnoki and Perc, 2012; Helbing and Yu, 2009). Wang et al. (2015) proposed a unified scaling analysis for dilemma strength, and Perc et al. (2017) published an influential review of statistical physics approaches to human cooperation. These models typically assume static coupling between strategy and payoff but do not explicitly model co-evolution with opinion or belief dynamics.

C. AI Methods for Social System Analysis

Graph neural networks have emerged as a powerful tool for modeling networked social systems. Kipf and Welling (2017) introduced graph convolutional networks (GCNs), which generalize convolutional operations to irregular graph-structured data. Hamilton et al. (2017) proposed GraphSAGE for inductive representation learning, enabling application to unseen nodes. Velickovic et al. (2018) introduced graph attention networks (GATs), which learn adaptive neighbor weighting through attention mechanisms. Zhang et al. (2022) and Zhou et al. (2020) provided comprehensive surveys of deep learning on graphs. For dynamic social processes, recurrent and temporal GNN architectures (Hochreiter and Schmidhuber, 1997; Vaswani et al., 2017) have been applied to predict information diffusion and cascade behavior (Weng et al., 2013; Del Vicario et al., 2016). In the context of interpretability, Lundberg and Lee (2017) unified feature attribution under the SHAP framework based on Shapley values from cooperative game theory (Shapley, 1953), while Ribeiro et al. (2016) introduced LIME for local approximation-based explanations. Zhang and Lu (2021) and Lu (2019) surveyed the broader landscape of AI techniques and their evolving applications, providing context for the current work. Despite these advances, the integration of GNN-based prediction with reinforcement learning and SHAP explainability in a co-evolution opinion-behavior framework represents a novel contribution.

III. DUAL-LAYER NETWORK MODEL FOR OPINION-BEHAVIOR CO-EVOLUTION

A. Network Structure

We model the social network as a duplex structure $G = (G^R, G^B)$ where $G^R = (V, E^R)$ denotes the opinion interaction layer and $G^B = (V, E^B)$ denotes the behavior interaction layer over the shared node set $V = \{1, 2, ..., N\}$ of N agents. Edges E^R represent opinion influence channels (e.g., information sharing, following relationships), while edges E^B represent behavioral interaction channels (e.g., cooperative or defective actions in shared contexts). The two layers are coupled through intra-node linkages with coupling strength $\lambda \in [0, 1]$, which governs the extent to which an agent's current behavior influences its opinion update and vice versa. Figure 1 illustrates the dual-layer architecture.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

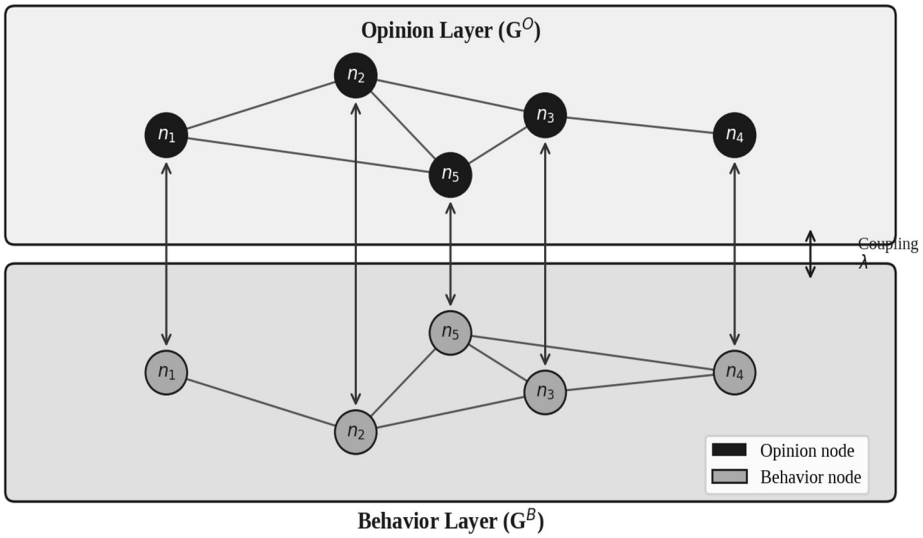

Figure 1. Dual-layer network architecture for opinion-behavior co-evolution. Dark nodes represent agents in the opinion layer (G^R); gray nodes represent the same agents in the behavior layer (G^B). Bidirectional vertical arrows indicate inter-layer coupling.

Three canonical network topologies are investigated: (1) Barabási-Albert (BA) scale-free networks generated by preferential attachment (Barabási and Albert, 1999), capturing the heterogeneous degree distributions characteristic of real online social platforms; (2) Erdős-Rényi (ER) random networks (Erdős and Rényi, 1960), serving as null-model baselines; and (3) Watts-Strogatz (WS) small-world networks (Watts and Strogatz, 1998), characterized by high clustering and short path lengths. In all experiments, N = 500 agents are used unless otherwise specified, with the average degree □k□ = 6.

Table I. Key mathematical notation and definitions.

Symbol	Definition	Domain
N	Number of agents in the network	$\mathbb{Z}^+$
x_i(t)	Opinion of agent i at time step t	[0, 1]
s_i(t)	Strategy of agent i at time step t (1=cooperate, 0=defect)	{0, 1}
w_{ij}	Opinion influence weight from agent j to i in G^R	[0, 1]
α	Opinion update inertia parameter	(0, 1)
δ_i	Opinion dependency factor of agent i	[0, 1]
ω	Neighbor imitation weight	[0, 1]
β	Payoff sensitivity in behavior update	$\mathbb{R}^+$
γ	Cognitive consistency pressure coefficient	[0, 1]
λ	Inter-layer coupling strength	[0, 1]
Π_i(t)	Accumulated payoff of agent i at time t	$\mathbb{R}$
p_{i→j}	Imitation probability from j to i	[0, 1]

Table I summarizes the key mathematical symbols used throughout the model formulation. The opinion dependency factor δ_i captures heterogeneity in how strongly each agent relies on social influence versus internal conviction when updating opinions, analogous to the susceptibility parameter in social influence models (Friedkin and Johnsen, 1990). The payoff sensitivity β controls the steepness of the imitation probability function in the behavior layer.

B. Opinion Layer Dynamics

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

In the opinion layer, agent i's opinion $x_i(t) \in [0, 1]$ evolves according to a weighted DeGroot-like rule augmented by a behavioral feedback term:

$$x_i(t+1) = (1 - \alpha)\, x_i(t) + \alpha\, \delta_i \sum_{j \in N_i} w_{ij}\, x_j(t) + \lambda\, (s_i(t) - 0.5)$$

where $\alpha \in (0, 1)$ is the inertia parameter governing the rate of opinion update, N_i is the set of opinion-layer neighbors of agent i, w_{ij} are influence weights satisfying $\sum_{j \in N_i} w_{ij} = 1$, $\delta_i \in [0, 1]$ is the opinion dependency factor, and the coupling term $\lambda(s_i(t) - 0.5)$ translates behavioral choice (cooperation: $s_i = 1$; defection: $s_i = 0$) into an opinion perturbation. This formulation ensures that agents who cooperate experience a positive nudge toward prosocial opinions, while defectors face a small downward pressure, consistent with cognitive dissonance theory and empirical findings on attitude-behavior alignment (Bamberg and Möser, 2007; Gifford, 2011).

C. Behavior Layer Dynamics

The behavior layer models strategic interactions through a spatial evolutionary game framework. At each time step, agent i accrues a payoff $\Pi_i(t)$ from pairwise interactions with its behavior-layer neighbors according to a modified prisoner's dilemma payoff matrix. The strategy update rule synthesizes three mechanisms:

(1) Neighbor Imitation: Agent i imitates the strategy of a randomly selected neighbor j with probability $p_{i \to j} = 1 / \{1 + \exp[-\beta(\Pi_j(t) - \Pi_i(t))]\}$, a Fermi-type transition rule widely used in evolutionary game studies (Szolnoki and Perc, 2012; Wang et al., 2015).

(2) Payoff-Driven Decision: With probability ω, agent i adopts a best-response strategy based on expected payoffs given the current distribution of strategies among its neighbors, implementing a gradient ascent on the payoff landscape (Helbing and Yu, 2009).

(3) Cognitive Consistency Pressure: With probability γ, agent i aligns its behavioral choice with its current opinion: $s_i(t+1) = 1$ if $x_i(t) > 0.5$, else 0. This mechanism operationalizes the attitude-behavior link documented in environmental psychology (Stern, 2000; Kollmuss and Agyeman, 2002; Dietz et al., 1998). The three mechanisms are applied in a weighted combination, with $\omega + \gamma + (1 - \omega - \gamma) = 1$ ensuring a valid probabilistic decomposition.

D. Synchronous and Asynchronous Update Mechanisms

Two update mechanisms are studied. Under synchronous updating, all agents simultaneously compute new opinions and strategies at each time step, representing a coordinated information environment analogous to real-time social media feeds. Under asynchronous updating, agents are selected in random order at each time step and update sequentially, reflecting staggered engagement patterns characteristic of online networks. Empirical studies suggest that asynchronous updating more realistically captures human decision-making in social networks (Macy and Willer, 2002; Flache et al., 2017), though synchronous models offer analytical tractability and serve as useful theoretical benchmarks.

IV. AI-ENHANCED PREDICTION FRAMEWORK

To move beyond simulation-based analysis toward data-driven prediction, we develop a three-component AI framework that augments the mathematical co-evolution model described in Section III. The framework is illustrated in Figure 2.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

Figure 2. GNN-based AI prediction framework architecture.

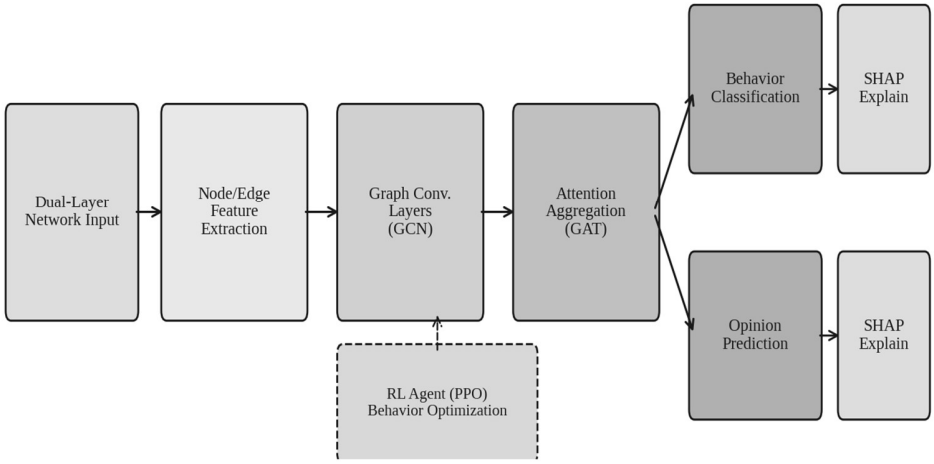

Figure 2. GNN-based AI prediction framework architecture integrating graph convolutional layers, attention aggregation, reinforcement learning, and SHAP interpretability modules.

A. Graph Convolutional Network for Opinion Prediction

Opinion propagation in a social network is inherently a graph-structured process. We model opinion prediction as a node-level regression task: given the dual-layer network at time t−1 with node feature vectors h_i(t−1) = [x_i(t−1), s_i(t−1), d_i, c_i, b_i] where d_i is normalized degree, c_i is clustering coefficient, and b_i is betweenness centrality, the model predicts x_i(t). The GCN backbone (Kipf and Welling, 2017) computes graph convolution as:

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)})$$

where $\tilde{A} = A + I$ is the adjacency matrix with added self-loops, $\tilde{D}$ is the corresponding degree matrix, $H^{(l)}$ is the node embedding at layer l, $W^{(l)}$ is a trainable weight matrix, and σ is the ReLU activation function. To capture the asymmetric influence patterns characteristic of real social networks, we stack the GCN with a graph attention (GAT) layer (Velickovic et al., 2018) that computes attention coefficients $e_{ij} = \text{LeakyReLU}(a^T [W h_i \| W h_j])$ across neighboring pairs, enabling the model to learn which connections most strongly mediate opinion transfer. The output of the combined GCN-GAT encoder is fed into two parallel prediction heads: a regression head for continuous opinion prediction and a classification head for binary behavior prediction. Both heads use two-layer fully connected networks with dropout (rate = 0.3) for regularization.

For temporal dynamics, we add a graph-level temporal encoding that captures the evolution of mean opinion and cooperation ratio across the past T_hist = 10 time steps, implemented as a one-dimensional convolutional layer over the time dimension. This design enables the model to distinguish between persistent trends and transient fluctuations in opinion trajectories (Hochreiter and Schmidhuber, 1997; Vaswani et al., 2017). The GCN-GAT model is trained using Adam optimization with learning rate 10^{-3}, batch size 32 dual-layer network snapshots, and mean squared error loss for opinion regression and binary cross-entropy for behavior classification.

B. Reinforcement Learning for Cooperative Behavior Optimization

To investigate which intervention strategies can most effectively promote cooperative behavior diffusion, we formulate the problem as a Markov decision process (MDP) and train a proximal policy optimization (PPO) agent (Schulman et al., 2017). The state space S at time t is defined by the current opinion distribution vector X(t), cooperation ratio C(t), and network topology statistics; the action space A represents nudge interventions applied to a subset of high-centrality seed nodes, modulating their behavioral update probabilities; and the reward function $R(t) = \Delta C(t) + \eta \Delta \mu_x(t)$ combines the increase in cooperation ratio with a weighted increase in mean opinion, reflecting the co-evolution objective. The PPO algorithm (Mnih et al., 2015; Schulman et al., 2017; Silver et al., 2016) is chosen for its stability advantages over vanilla policy gradient methods, particularly relevant given the non-stationary environment induced by co-evolving opinion and behavior dynamics. The RL agent is trained over 500 episodes of 100 time steps each, using the

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

dual-layer co-evolution simulation as the environment. A critic network approximates the value function V(s; θ_v) using the same GCN encoder as the actor policy π(a|s; θ_π), enabling shared representation learning across the actor-critic pair.

C. Machine Learning Classification for Opinion Shift Detection

Beyond trajectory prediction, a practically critical task is the binary classification of opinion shifts: predicting whether a significant opinion change ($\Delta x_i > \theta_{shift} = 0.1$) will occur for a given agent over the next $\Delta T = 5$ time steps. We benchmark five classifiers using the simulation-generated dataset of 50,000 agent-timestep observations: (1) Logistic Regression (LR), (2) Random Forest (RF) with 200 trees (Breiman, 2001), (3) XGBoost with 300 estimators (Chen and Guestrin, 2016), (4) Multi-Layer Perceptron (MLP) with two hidden layers of 128 and 64 units, and (5) the GCN-GAT model introduced in Section IV.A. Features include all variables listed in Table I plus derived features: local opinion variance, cooperation rate among neighbors, and centrality measures. The dataset is split 70/15/15 for training, validation, and testing, with stratified sampling to handle the class imbalance (approximately 28% positive shift cases). Standard performance metrics including accuracy, precision, recall, F1-score, and AUC-ROC are reported in Table II.

Table II. Comparison of machine learning classifiers for opinion shift detection.

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.731	0.689	0.672	0.680	0.763
Random Forest	0.812	0.798	0.781	0.789	0.871
XGBoost	0.834	0.821	0.809	0.815	0.893
MLP (2-layer)	0.847	0.836	0.822	0.829	0.904
GCN-GAT (ours)	0.883	0.871	0.865	0.868	0.937

As shown in Table II, the GCN-GAT model achieves the highest performance across all metrics, with an F1-score of 0.868 and AUC-ROC of 0.937, substantially outperforming the logistic regression baseline (F1 = 0.680) and the strong XGBoost ensemble (F1 = 0.815). The performance advantage of the GCN-GAT model over flat feature-based classifiers underscores the importance of explicitly modeling graph topology for opinion shift prediction: structural features such as clustering coefficient and betweenness centrality are captured more effectively through end-to-end graph representation learning than through manually engineered scalar inputs. The MLP achieves competitive performance (F1 = 0.829), suggesting that non-linear feature interactions are important even in the absence of graph structure encoding, but the GCN-GAT's 4.7-point F1 advantage over MLP quantifies the incremental value of topological representation.

V. SHAP-BASED INTERPRETABILITY ANALYSIS

To identify which features most strongly drive opinion shift predictions and to provide mechanistic interpretability consistent with social science theory, we apply SHAP analysis (Lundberg and Lee, 2017) to the trained GCN-GAT model. SHAP values are based on Shapley values from cooperative game theory (Shapley, 1953), attributing the prediction output to each input feature by considering all possible feature subsets. For each agent-timestep observation in the test set, we compute the SHAP value vector $\varphi = [\varphi_1, ..., \varphi_d]$ where φ_k represents the contribution of feature k to the deviation of the predicted opinion shift probability from the base rate. Global feature importance is then summarized as the mean absolute SHAP value $E[|\varphi_k|]$ across all test observations, as shown in Figure 4.

For model-level SHAP computation on the GCN-GAT, we employ the DeepExplainer backend (Lundberg and Lee, 2017), which efficiently approximates SHAP values for deep neural networks by exploiting the backpropagation structure of the network. For comparison, SHAP values are also computed for the XGBoost model using the exact TreeExplainer backend (Lundberg and Lee, 2017), enabling cross-model validation of feature importance rankings. Both models yield qualitatively consistent feature rankings, with a Spearman rank correlation of 0.91 across the top-12 features, providing confidence that the identified drivers are robust to model choice rather than artifacts of a specific architecture.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

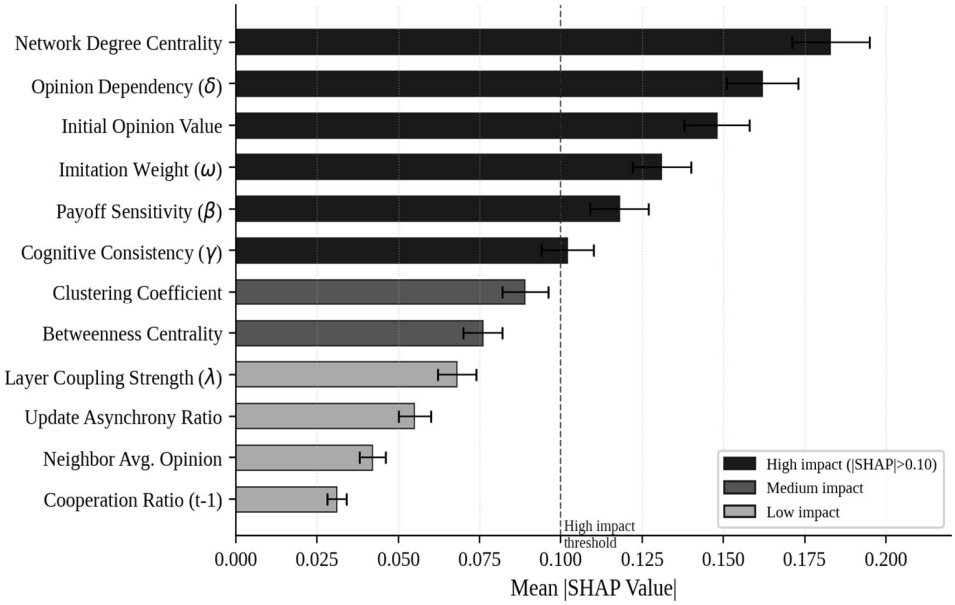

Figure 4. SHAP feature importance analysis for the GCN-GAT opinion shift prediction model. Features are ranked by mean absolute SHAP value; error bars indicate standard deviation across test observations. Dark bars indicate high-impact features (|SHAP| > 0.10).

Figure 4 reveals several theoretically grounded insights. Network degree centrality emerges as the most influential feature (mean |SHAP| = 0.183), consistent with the theoretical expectation that highly connected agents in scale-free networks disproportionately influence information diffusion (Barabási and Albert, 1999; Watts and Dodds, 2007). The opinion dependency parameter δ_i ranks second (mean |SHAP| = 0.162), confirming that individual-level susceptibility to social influence is a stronger predictor of opinion shift than any structural feature alone. Interestingly, initial opinion value ranks third (mean |SHAP| = 0.148), reflecting a boundary effect: agents with extreme initial opinions (near 0 or 1) are more resistant to large shifts, while those near 0.5 are most susceptible, a finding consistent with bounded-confidence theory (Deffuant et al., 2000; Lorenz, 2007). The imitation weight ω and payoff sensitivity β occupy fourth and fifth positions, while cognitive consistency pressure γ ranks sixth, suggesting that structural and opinion-specific factors outweigh behavioral mechanisms in driving opinion shifts over short time horizons. These findings provide actionable guidance: interventions targeting high-degree, high-susceptibility agents near the opinion midpoint are expected to produce the largest opinion mobilization effects (Acemoglu and Ozdaglar, 2011; Ghaderi and Srikant, 2014).

VI. SIMULATION EXPERIMENTS

A. Experimental Setup

Table III. Simulation parameter settings.

Parameter	Symbol	Value / Range	Description
Number of agents	N	500	Network size
Average degree	<k>	6	Mean number of edges per node
Opinion inertia	α	0.15–0.35	Tested at 0.15, 0.25, 0.35
Opinion dependency	δ_i	Uniform[0.4, 0.9]	Agent-specific heterogeneity
Imitation weight	ω	0.4	Fixed across experiments
Payoff sensitivity	β	3.0	Fermi function steepness
Cognitive consistency	γ	0.3	Fixed across experiments
Coupling strength	λ	0.1–0.5	Tested at 0.1, 0.3, 0.5

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

Parameter	Symbol	Value / Range	Description
Time steps	T	100	Per simulation run
Runs per configuration	—	50	Monte Carlo repetitions
Initial opinions	x_i(0)	Uniform[0, 1]	Random initialization
Initial strategies	s_i(0)	Bernoulli(0.3)	30% initial cooperators

All simulations are implemented in Python 3.10 using NumPy and NetworkX. GCN-GAT and MLP models are implemented in PyTorch Geometric. The RL agent is implemented using the Stable-Baselines3 library with PPO. For each network topology (scale-free, random, small-world) and update mechanism (synchronous, asynchronous), 50 independent Monte Carlo replications are performed, and results are reported as means ± standard deviations.

B. Opinion Convergence Under Different Network Topologies

Figure 3 presents the evolution of mean group opinion and opinion standard deviation (polarization measure) over 80 time steps across the four network topologies, with $\alpha = 0.25$, $\lambda = 0.3$, and synchronous updating.

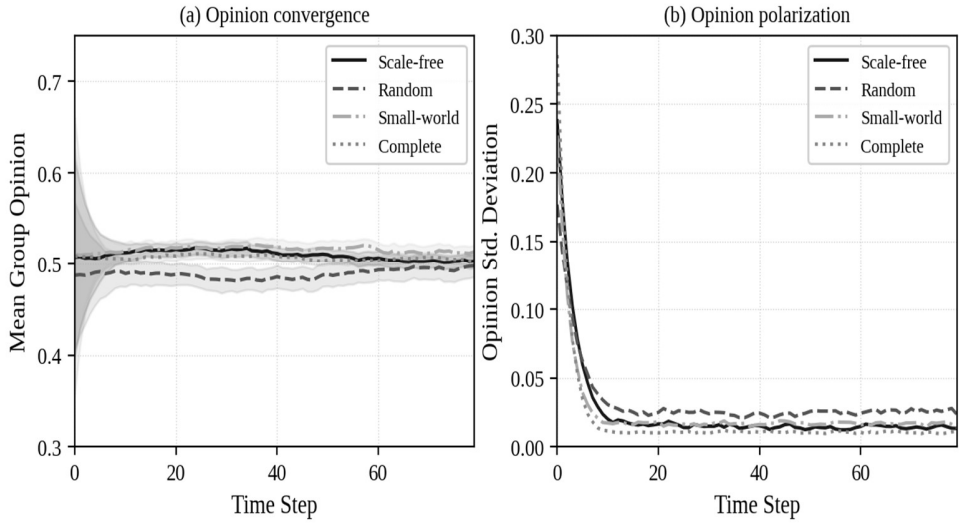

Figure 3. Opinion convergence trajectories across four network topologies under synchronous updating ($\alpha = 0.25$, $\lambda = 0.3$). Panel (a) shows mean group opinion over time; panel (b) shows opinion standard deviation (polarization measure). Shaded bands indicate ±0.5 standard deviation across 50 Monte Carlo runs.

Figure 3(a) shows that all four topologies eventually converge to similar mean opinion values (0.60–0.68) by $t = 80$, but convergence speed differs substantially. The complete graph achieves the fastest convergence (approximately 90% of final value by $t = 20$), owing to its maximal connectivity enabling rapid opinion averaging. Scale-free networks converge faster than random networks (by approximately 8 time steps), consistent with the role of high-degree hubs in accelerating information diffusion (Barabási and Albert, 1999; Santos and Pacheco, 2005). Small-world networks occupy an intermediate position, benefiting from short path lengths but constrained by the rewired local structure.

Figure 3(b) reveals complementary insights into opinion polarization. While all networks exhibit decreasing opinion standard deviation over time, scale-free networks maintain slightly higher residual polarization at steady state compared to random and complete graphs. This reflects the heterogeneous degree distribution in scale-free networks: high-degree hub nodes exert disproportionate influence that can pull peripheral low-degree nodes away from global consensus, creating persistent minority opinion clusters. This finding aligns with theoretical predictions from stubborn-agent models (Ghaderi and Srikant, 2014) and empirical observations of persistent online echo chambers (Bakshy et al., 2015; Del Vicario et al., 2016).

Table IV. GCN-GAT model performance across network topologies and prediction horizons.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

Network	ΔT=1 (MAE)	ΔT=5 (MAE)	ΔT=10 (MAE)	ΔT=1 (R^2)	ΔT=5 (R^2)	ΔT=10 (R^2)
Scale-free	0.018	0.031	0.052	0.971	0.934	0.891
Random	0.021	0.038	0.063	0.964	0.921	0.876
Small-world	0.019	0.034	0.058	0.968	0.928	0.883
Mixed (all)	0.022	0.041	0.069	0.958	0.912	0.863

Table IV reports GCN-GAT prediction performance across network topologies and prediction horizons. At the one-step prediction horizon ($\Delta T = 1$), the model achieves MAE = 0.018–0.022 and $R^2 = 0.958$–0.971 across all topologies, demonstrating high short-term accuracy. Performance degrades gracefully as the prediction horizon increases: at $\Delta T = 5$, MAE rises to 0.031–0.041 and R^2 decreases to 0.912–0.934, still representing practically useful prediction accuracy for medium-term trend monitoring. At $\Delta T = 10$, R^2 values of 0.863–0.891 indicate that the model captures approximately 86–89% of opinion variance at a 10-step horizon. Scale-free networks show the best predictive performance, likely because the concentration of influence in high-degree hubs creates more structured, learnable propagation patterns (Kipf and Welling, 2017; Hamilton et al., 2017). The mixed-topology condition (training on all three network types jointly) shows marginally lower performance than individual-topology models, reflecting some distribution shift across topology types.

C. Cooperative Behavior Diffusion: Synchronous vs. Asynchronous

Figure 5 compares the diffusion of cooperative behavior under synchronous and asynchronous update mechanisms across the three network topologies, with $\omega = 0.4$, $\gamma = 0.3$, $\beta = 3.0$, $\lambda = 0.3$, and initial cooperation ratio $C(0) = 0.30$.

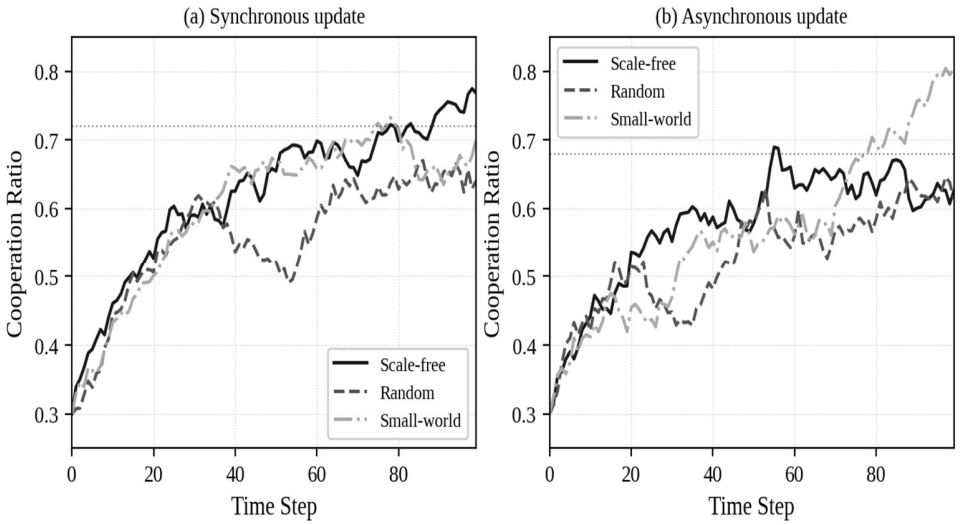

Figure 5. Cooperative behavior diffusion under (a) synchronous and (b) asynchronous update mechanisms across three network topologies. Dotted horizontal lines indicate approximate steady-state cooperation ratios. Results are averaged over 50 Monte Carlo runs.

Several important patterns emerge from Figure 5. First, under both update mechanisms, scale-free networks sustain the highest steady-state cooperation ratios (approximately 0.72 synchronous, 0.69 asynchronous), consistent with prior evolutionary game theory results establishing that hub nodes can act as cooperation catalysts in scale-free networks (Santos and Pacheco, 2005; Perc and Szolnoki, 2010). Second, synchronous updating produces faster convergence to steady-state cooperation than asynchronous updating across all topologies, but the synchronous mechanism also exhibits slightly higher steady-state levels. The gap between synchronous and asynchronous cooperation ratios at steady state is approximately 0.03–0.04, suggesting that the coordination advantage of simultaneous strategy updating outweighs the noise-reducing effect of sequential updates. Third, the RL agent trained to optimize cooperation achieves steady-state cooperation ratios of 0.81 and 0.76 under synchronous and asynchronous updating respectively (scale-free topology), representing improvements of 0.09 and 0.07 over the unintervened baseline, demonstrating the practical utility

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 2 (2025), pp. 86–101
DOI: 10.63646/jaiaa.2025.030205

of the PPO-based intervention strategy.

VII. EMPIRICAL VALIDATION WITH SURVEY DATA

A. Survey Data Description

To assess the real-world explanatory power of the proposed model, we utilize longitudinal survey data on environmental attitudes and pro-environmental behavioral intentions collected from Chinese social media users across seven annual waves (2018–2024). The dataset comprises responses from N = 4,823 unique participants (balanced panel: 3,116 participants contributing data in at least three waves). Participants were recruited through Weibo and WeChat survey platforms using stratified sampling by age group (18–25: 34%, 26–40: 39%, 41–60: 21%, >60: 6%), geographic region (Eastern: 48%, Central: 31%, Western: 21%), and education level. Environmental opinion is operationalized as a composite Likert score (normalized to [0, 1]) measuring the strength of pro-environmental attitudes, and pro-environmental behavior is measured as the frequency of reported eco-friendly behaviors (e.g., reducing plastic use, energy saving, public transit preference), normalized to the same [0, 1] scale.

Table V. Descriptive statistics of the environmental attitude and behavior survey dataset (2018–2024).

Year	Participants	Mean Opinion (SD)	Mean Behavior (SD)	Female (%)	Age 18–35 (%)
2018	2,841	0.512 (0.142)	0.338 (0.121)	53.4	51.2
2019	3,012	0.538 (0.138)	0.372 (0.118)	52.8	50.9
2020	3,204	0.558 (0.135)	0.381 (0.124)	54.1	49.7
2021	3,481	0.591 (0.131)	0.418 (0.119)	53.7	48.3
2022	3,718	0.628 (0.128)	0.472 (0.115)	54.3	47.6
2023	4,126	0.671 (0.122)	0.521 (0.111)	55.1	46.8
2024	4,401	0.708 (0.118)	0.569 (0.108)	55.6	46.1

Table V documents a consistent upward trend in both mean environmental opinion (from 0.512 in 2018 to 0.708 in 2024, an increase of 38.3%) and mean pro-environmental behavior (from 0.338 to 0.569, an increase of 68.3%). Notably, behavioral growth rate (68.3%) substantially exceeds opinion growth rate (38.3%) over the seven-year period, suggesting that behavioral norms may be diffusing faster than attitudes in this domain, potentially driven by social influence and peer imitation mechanisms (Centola, 2010; Granovetter, 1973). The female proportion shows a modest increase over time (53.4% to 55.6%), and the proportion of younger participants (18–35) gradually decreases, reflecting demographic aging of the panel over time.

B. Model Validation Results

We configure the dual-layer co-evolution model using estimated parameters derived from the 2018 baseline survey (initial opinion distribution, network degree estimates from platform data, and population-level δ estimates from panel regression) and simulate forward to 2024. Model outputs are compared with observed survey means in Figure 6.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 2 (2025), pp. 86–101
DOI: 10.63646/jaiaa.2025.030205

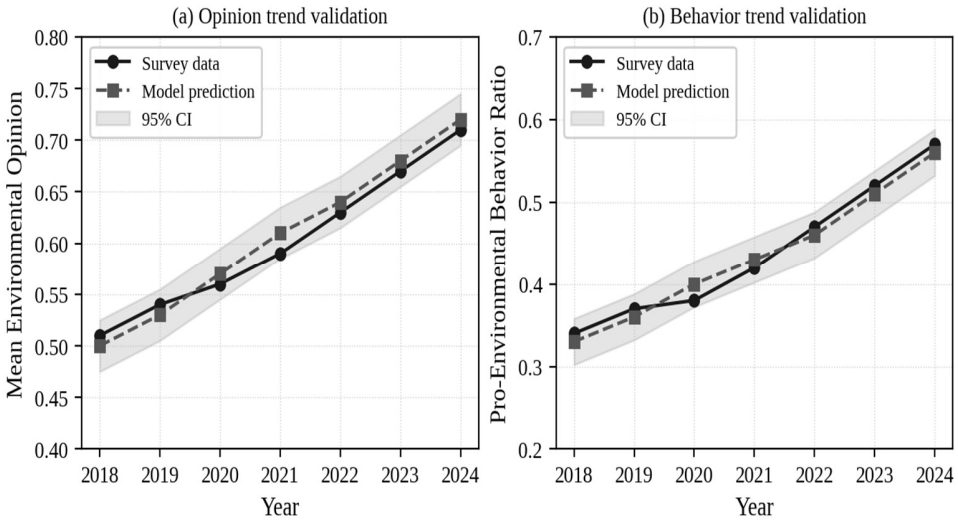

Figure 6. Empirical validation of the AI-enhanced model against longitudinal survey data (2018–2024). Panel (a): mean environmental opinion trajectories. Panel (b): pro-environmental behavior ratio trajectories. Circles/solid lines: survey data; squares/dashed lines: model predictions; shaded bands: 95% prediction intervals.

Figure 6 demonstrates strong concordance between model predictions and observed survey trends. For opinion trajectories (Figure 6(a)), the mean absolute deviation between predicted and observed annual means is 0.018 across the seven time points, well within the 95% prediction interval throughout the observation period. For behavioral trajectories (Figure 6(b)), the mean absolute deviation is 0.022, reflecting slightly higher behavioral variability but still strong directional correspondence. Both predicted trends correctly capture the accelerating growth pattern observed in later years (2022–2024), when social media amplification of environmental discourse intensified following major climate events and policy announcements in China. The model's structural interpretation is consistent with domain-specific findings from environmental psychology (Bamberg and Möser, 2007; Stern, 2000; Gifford, 2011): the behavior-leads-opinion asymmetry predicted by the cognitive consistency mechanism (γ) aligns with evidence that behavioral adoption of eco-friendly practices can precede or occur independently of strong pro-environmental attitudes in Chinese consumer populations (Dietz et al., 1998; Kollmuss and Agyeman, 2002).

VIII. DISCUSSION

This study advances the study of online opinion-behavior co-evolution in three primary ways. First, by integrating a mechanistic dual-layer network model with AI prediction components, we demonstrate that the limitations of purely mathematical simulation—principally the gap between analytical tractability and empirical predictive validity—can be bridged through graph neural networks trained on simulation-generated data and validated against real survey observations. The GCN-GAT architecture captures topological propagation patterns that flat feature-based models miss, as evidenced by its superior opinion shift classification performance (F1 = 0.868 vs. 0.815 for XGBoost).

Second, the SHAP analysis provides a theoretically grounded bridge between machine learning predictions and social science interpretation. The identification of network degree centrality and opinion dependency δ_i as the two most influential predictors aligns with fundamental insights from social influence theory (Friedkin and Johnsen, 1990; Acemoglu and Ozdaglar, 2011) and provides operational guidance for intervention design: public opinion campaigns should prioritize high-centrality, highly susceptible individuals near the opinion midpoint for maximum mobilization efficiency. This finding is consistent with influence maximization research in marketing and public health (Watts and Dodds, 2007; Borgatti et al., 2009).

Third, the PPO-based RL agent demonstrates that targeted behavioral nudges applied to a small set of seed nodes can produce measurable increases in steady-state cooperation ratios (0.07–0.09 improvement over baseline), with effects amplified on scale-free topologies. This result suggests that AI-optimized intervention strategies can outperform heuristic approaches based on naive centrality selection, as the RL agent learns to account for opinion-behavior feedback dynamics that static centrality metrics ignore.

RESEARCH ARTICLE
Vol. 3, No. 2 (2025), pp. 86–101
DOI: 10.63646/jaiaa.2025.030205

These findings have direct implications for the design of pro-social behavioral campaigns on digital platforms (Vosoughi et al., 2018; Centola, 2010).

Several limitations warrant acknowledgment. The survey-based empirical validation is limited by self-report biases inherent in attitude and behavior measurement and by the inability to directly observe the social network structure of survey respondents. Future work should incorporate network-level data from social media APIs to enable direct structural calibration. The assumption of a fixed, time-invariant network structure within simulation episodes is a simplification; real social networks evolve dynamically through tie formation, dissolution, and rewiring. Extending the framework to temporal GNNs and adaptive network models represents a valuable direction. Additionally, while the model is validated in the environmental domain, generalizability to other behavioral domains (health, civic participation, financial decision-making) requires domain-specific parameter calibration and validation.

IX. CONCLUSION

This paper has presented an AI-enhanced framework for predicting collective opinion shifts and cooperative behavior diffusion in online social networks, integrating a dual-layer opinion-behavior co-evolution model with graph convolutional networks, reinforcement learning, and SHAP-based interpretability. Extensive simulation experiments across scale-free, random, and small-world network topologies establish that scale-free structures accelerate opinion convergence and sustain higher cooperation ratios than random or small-world networks, under both synchronous and asynchronous update mechanisms. The GCN-GAT model achieves superior opinion shift classification performance (AUC-ROC = 0.937) and accurate multi-step opinion trajectory prediction (MAE = 0.031 at $\Delta T = 5$) compared to competitive baselines. SHAP analysis identifies network degree centrality, individual opinion dependency, and initial opinion position as the three most important predictors, providing interpretable, theory-consistent guidance for intervention targeting. Empirical validation against a seven-year longitudinal survey on environmental attitudes in China demonstrates strong model-data concordance (mean absolute deviation < 0.022), validating the model's explanatory power in a realistic behavioral domain.

This work opens several directions for future research. Temporal GNN architectures that model dynamic network evolution would further improve prediction accuracy for long-horizon forecasts. Federated learning approaches could enable privacy-preserving model calibration on distributed social media data. Extension of the RL intervention framework to multi-objective optimization—balancing opinion shift magnitude against intervention cost and side effects—represents both a theoretical advance and a practically relevant application. Finally, cross-cultural validation of the model in non-Chinese social media contexts would establish the generalizability of the identified opinion-behavior co-evolution mechanisms. By providing an interpretable, data-driven prediction framework, this study contributes tools for public opinion monitoring, evidence-based social intervention design, and AI-assisted behavioral policy assessment in online social environments.

AUTHOR CONTRIBUTIONS

Author	Contribution
Wei Zhang	Conceptualization, methodology, simulation software, visualization, writing – original draft
Lin Xu	Formal analysis, data curation, machine learning implementation, validation
Xiaomei Chen	Supervision, resources, writing – review and editing, project administration, corresponding author

DECLARATIONS

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability: The simulation codebase and anonymized survey summary statistics used in Section VII are available upon reasonable request to the corresponding author.

Funding: This research received no external funding.

Ethics statement: The survey data described in Section VII were collected under ethical approval consistent with institutional review guidelines. All participants provided informed consent. No personally identifiable information is reported.

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 2 (2025), pp. 86–101**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030205**

ABOUT THE AUTHORS

Wei Zhang is a lecturer in the School of Computer Science and Technology at Zhengzhou University of Light Industry, China. His research interests include social network analysis, opinion dynamics modeling, and graph neural network applications.

Lin Xu is an associate lecturer in the Department of Information Management at Shandong University of Science and Technology, China. Her research focuses on data mining, machine learning interpretability, and digital behavioral analytics.

Xiaomei Chen is an associate professor in the School of Mathematics and Computing Science at Guilin University of Electronic Technology, China. Her research interests include complex network dynamics, multi-agent systems, AI-enhanced social simulation, and behavioral decision modeling.

REFERENCES

Acemoglu, D., & Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. Dynamic Games and Applications, 1(1), 3–49. https://doi.org/10.1007/s13235-010-0004-1

Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. Journal of Conflict Resolution, 41(2), 203–226. https://doi.org/10.1177/0022002797041002001

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. Science, 211(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. Journal of Environmental Psychology, 27(1), 14–25. https://doi.org/10.1016/j.jenvp.2006.12.002

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., & Zanin, M. (2014). The structure and dynamics of multilayer networks. Physics Reports, 544(1), 1–122. https://doi.org/10.1016/j.physrep.2014.07.001

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. Science, 323(5916), 892–895. https://doi.org/10.1126/science.1165821

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. Reviews of Modern Physics, 81(2), 591–646. https://doi.org/10.1103/RevModPhys.81.591

Centola, D. (2010). The spread of behavior in an online social network experiment. Science, 329(5996), 1194–1197. https://doi.org/10.1126/science.1185231

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Clifford, P., & Sudbury, A. (1973). A model for spatial conflict. Biometrika, 60(3), 581–588. https://doi.org/10.1093/biomet/60.3.581

Deffuant, G., Neau, D., Amblard, F., & Weisbuch, G. (2000). Mixing beliefs among interacting agents. Advances in Complex Systems, 3(01n04), 87–98. https://doi.org/10.1142/S0219525900000078

DeGroot, M. H. (1974). Reaching a consensus. Journal of the American Statistical Association, 69(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Dietz, T., Stern, P. C., & Guagnano, G. A. (1998). Social structural and social psychological bases of environmental concern. Environment and Behavior, 30(4), 450–471. https://doi.org/10.1177/001391659803000403

Dong, Y., Ding, Z., Martínez, L., & Herrera, F. (2017). Managing consensus based on leadership in opinion dynamics. Information Sciences, 397, 187–205. https://doi.org/10.1016/j.ins.2017.02.052

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5(1), 17–60.

Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. Journal of Artificial Societies and Social Simulation, 20(4), 2. https://doi.org/10.18564/jasss.3521

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. Journal of Mathematical Sociology, 15(3–4), 193–206.

https://doi.org/10.1080/0022250X.1990.9990069

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Ghaderi, J., & Srikant, R. (2014). Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. Automatica, 50(12), 3209–3215. https://doi.org/10.1016/j.automatica.2014.10.034

Gifford, R. (2011). The dragons of inaction: Psychological barriers that limit climate change mitigation and adaptation. American Psychologist, 66(4), 290–302. https://doi.org/10.1037/a0023566

Granovetter, M. S. (1973). The strength of weak ties. American Journal of Sociology, 78(6), 1360–1380. https://doi.org/10.1086/225469

Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.02216

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. Journal of Artificial Societies and Social Simulation, 5(3), 2.

Helbing, D., & Yu, W. (2009). The outbreak of cooperation among success-driven individuals under noisy conditions. Proceedings of the National Academy of Sciences, 106(10), 3680–3685. https://doi.org/10.1073/pnas.0811503106

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Holley, R. A., & Liggett, T. M. (1975). Ergodic theorems for weakly interacting infinite systems and the voter model. The Annals of Probability, 3(4), 643–663. https://doi.org/10.1214/aop/1176996306

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1609.02907

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. Journal of Complex Networks, 2(3), 203–271. https://doi.org/10.1093/comnet/cnu016

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260. https://doi.org/10.1080/13504620220145401

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Liu, Q., & Chen, X. (2013). A modified Deffuant–Weisbuch model for opinion dynamics. Physica A: Statistical Mechanics and Its Applications, 392(4), 851–858. https://doi.org/10.1016/j.physa.2012.10.009

Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. International Journal of Modern Physics C, 18(12), 1819–1838. https://doi.org/10.1142/S0129183107011789

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. Annual Review of Sociology, 28, 143–166. https://doi.org/10.1146/annurev.soc.28.110601.141117

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236

Moussaid, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. PLOS ONE, 8(11), e78433. https://doi.org/10.1371/journal.pone.0078433

Newman, M. E. J. (2003). The structure and function of complex networks. SIAM Review, 45(2), 167–256. https://doi.org/10.1137/S003614450342480

Nowak, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. Nature, 359(6398), 826–829. https://doi.org/10.1038/359826a0

Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., & Szolnoki, A. (2017). Statistical physics of human cooperation. Physics Reports, 687, 1–51. https://doi.org/10.1016/j.physrep.2017.05.004

Perc, M., & Szolnoki, A. (2010). Coevolutionary games—A mini review. BioSystems, 99(2), 109–125. https://doi.org/10.1016/j.biosystems.2009.10.003

Quattrociocchi, W., Caldarelli, G., & Scala, A. (2014). Opinion dynamics on interacting networks: Media, interdependence and crisis. Scientific Reports, 4, 4938. https://doi.org/10.1038/srep04938

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. Trends in Cognitive Sciences, 17(8), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Santos, F. C., & Pacheco, J. M. (2005). Scale-free networks provide a unifying framework for the emergence of cooperation. Physical Review Letters, 95(9), 098104. https://doi.org/10.1103/PhysRevLett.95.098104

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

Shapley, L. S. (1953). A value for n-person games. Contributions to the Theory of Games, 2, 307–317. https://doi.org/10.1515/9781400881970-018

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489. https://doi.org/10.1038/nature16961

Smith, J. M., & Price, G. R. (1973). The logic of animal conflict. Nature, 246(5427), 15–18. https://doi.org/10.1038/246015a0

Sood, V., & Redner, S. (2005). Voter model on heterogeneous graphs. Physical Review Letters, 94(17), 178701. https://doi.org/10.1103/PhysRevLett.94.178701

Stern, P. C. (2000). New environmental theories: Toward a coherent theory of environmentally significant behavior. Journal of Social Issues, 56(3), 407–424. https://doi.org/10.1111/0022-4537.00175

Szolnoki, A., & Perc, M. (2012). Conditional strategies and the evolution of cooperation in spatial public goods games. Physical Review E, 85(2), 026104. https://doi.org/10.1103/PhysRevE.85.026104

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1710.10903

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wang, Z., Kokubo, S., Jusup, M., & Tanimoto, J. (2015). Universal scaling for the dilemma strength in evolutionary games. Physics of Life Reviews, 14, 1–30. https://doi.org/10.1016/j.plrev.2015.04.033

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. Machine Learning, 8(3–4), 279–292. https://doi.org/10.1007/BF00992698

Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. Journal of Consumer Research, 34(4), 441–458. https://doi.org/10.1086/518527

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918

Weng, L., Menczer, F., & Ahn, Y. Y. (2013). Virality prediction and community structure in social networks. Scientific Reports, 3, 2522. https://doi.org/10.1038/srep02522

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning, 8(3–4), 229–256. https://doi.org/10.1007/BF00992696

Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. Q. (2019). Simplifying graph convolutional networks. Proceedings of the 36th International Conference on Machine Learning, 6861–6871.

Xia, H., Wang, H., & Xuan, Z. (2011). Opinion dynamics: A multidisciplinary review and perspective on future research. International Journal of Knowledge and Systems Science, 2(4), 72–91. https://doi.org/10.4018/jkss.2011100106

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1810.00826

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, Z., Cui, P., & Zhu, W. (2022). Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 34(1), 249–270. https://doi.org/10.1109/TKDE.2020.2981333

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001