

Unicoon: Hypergraph-based Multi-Agent Simulation of Information Cocoons

Chunyu Wei*
Renmin University of China
Beijing, China
weicy15@icloud.com

Yongsiqu Tu*
Renmin University of China
Beijing, China
tuyongsiqi@mails.neu.edu.cn

Yunhai Wang†
Renmin University of China
Beijing, China
cloudseawang@gmail.com

Abstract

Information cocoons pose significant challenges to democratic discourse and social cohesion. While extensive research has examined how social networks and recommender systems independently contribute to this phenomenon, their coevolving dynamics remain unexplored. We present Unicoon, the first unified computational framework that simultaneously models both social network propagation and algorithmic content delivery using LLM-based multi-agent simulation. Our key innovation lies in a hypergraph formulation where agents constitute nodes, social relationships form edges, and recommender-delivered content creates dynamic hyperedges, elegantly capturing heterogeneous information diffusion patterns within a single mathematical structure. To address the adaptive nature of recommendation algorithms, we introduce a dynamic hyperedge construction technique that computationally matches trending content with interested user cohorts in real-time. Extensive experiments on synthetic and real-world networks reveal that the synergistic interplay of social and algorithmic mechanisms creates qualitatively distinct polarization patterns, with a critical finding that larger macro-structures paradoxically accelerate micro-level cohort polarization.¹

CCS Concepts

• Human-centered computing → Social networks; • Computing methodologies → Multi-agent systems.

Keywords

Multi-agent Simulation, Large Language Models, Social Network

ACM Reference Format:

Chunyu Wei, Yongsiqu Tu, and Yunhai Wang. 2026. Unicoon: Hypergraph-based Multi-Agent Simulation of Information Cocoons. In *Proceedings of the ACM Web Conference 2026 (WWW '26)*, April 13–17, 2026, Dubai, United Arab Emirates. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3774904.3792486>

*These authors contributed equally to this work.

†Corresponding author.

¹Code is available at <https://github.com/Unicoon-simulation/Unicoon>

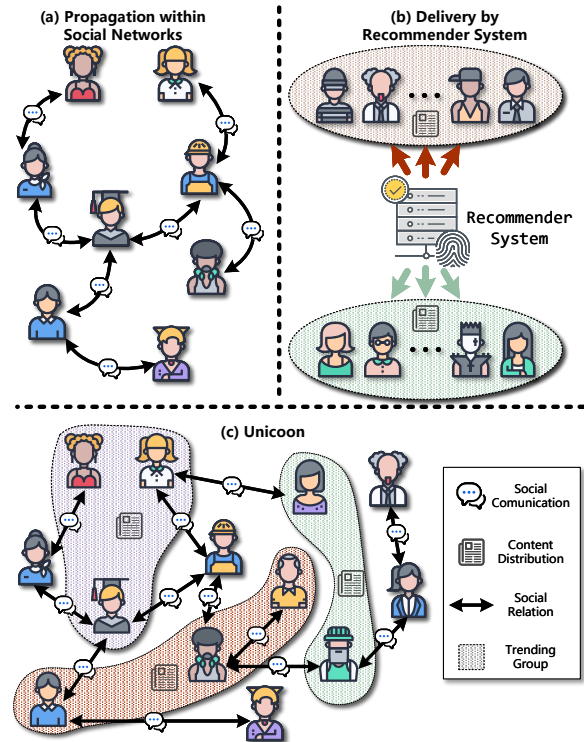


Figure 1: Information diffusion mechanisms in isolation and unified. (a) Social networks enable pairwise propagation through connections. (b) Recommender systems deliver content to groups with similar preferences. (c) Our Unicoon framework unifies mechanisms via a hypergraph: edges represent social relations while hyperedges capture algorithmic content distribution to trending groups.

1 Introduction

The pervasive proliferation of digital platforms and algorithmic recommendation systems has fundamentally transformed how individuals access and consume information in contemporary society. While these technological advances have enhanced user experience by delivering personalized content, they have simultaneously precipitated the emergence of information cocoons—isolated information environments where individuals are primarily exposed to content that reinforces their existing beliefs and preferences. This phenomenon has garnered significant scholarly attention due to its profound implications for opinion polarization, democratic discourse, and social cohesion.



This work is licensed under a Creative Commons Attribution 4.0 International License. *WWW '26, Dubai, United Arab Emirates*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2307-0/2026/04
<https://doi.org/10.1145/3774904.3792486>

Recent research has increasingly leveraged large language model (LLM)-based multi-agent simulation to investigate information cocoon dynamics, recognizing the potential of computational approaches to model complex social phenomena. However, existing studies exhibit a fundamental limitation: they examine the contributing factors to information cocoon formation in isolation. Specifically, current research predominantly focuses on either (1) **information propagation patterns within social networks**, analyzing network dynamics and echo chamber effects through pairwise interactions, or (2) **the influence of recommender systems**, studying how algorithmic content exposure shapes individual beliefs through group-wise information delivery [72, 79, 80].

This bifurcated approach overlooks a critical insight from recent empirical studies: information cocoons emerge from the coevolving and synergistic effects of social networks and recommender systems operating simultaneously. As illustrated in Figure 1, social networks facilitate information diffusion through interpersonal relationships and trust-based connections (Figure 1a), while recommender systems algorithmically curate and distribute content based on inferred user preferences (Figure 1b). The interplay between these two mechanisms creates complex feedback loops that amplify information cocoon formation beyond what either mechanism could achieve independently.

Despite this understanding, no existing study has employed LLM-based multi-agent systems to model these dual effects concurrently within a unified computational framework. This represents a significant gap in our ability to comprehensively understand and potentially mitigate information cocoon phenomena.

To address this limitation, we propose a novel unified simulation framework called **Unicoon** (Figure 1c) that simultaneously models both social network and recommender system effects on information cocoon formation. This endeavor requires overcoming two fundamental technical challenges:

- **Heterogeneous Diffusion Patterns.** Social networks and recommender systems exhibit fundamentally different information propagation mechanisms. Social networks operate through **pair-wise information transfer**, where information flows between directly connected individuals based on social relationships and trust. In contrast, recommender systems employ **group-wise information transfer**, algorithmically delivering content to cohorts of users with similar inferred preferences, regardless of their social connections.
- **Dynamic Recommendation Mechanisms.** While social networks possess relatively static structural properties that govern information flow, recommender systems exhibit dynamic and content-dependent targeting. For instance, sports-related content about swimming competitions will be delivered to users interested in aquatic sports, while ping-pong match coverage targets table tennis enthusiasts. This creates an adaptive information distribution mechanism where the target audience dynamically shifts based on content characteristics.

To handle the heterogeneous information diffusion patterns, we introduce an innovative hypergraph formulation where agents represent nodes, social relationships constitute traditional edges, and trending information from recommender systems forms hyperedges. This elegant data structure naturally captures both pairwise

social interactions and group-wise algorithmic content delivery within a single unified structure, as visualized in Figure 1c.

To address the dynamic nature of recommender systems, we develop a dynamic hyperedge construction technique that computationally simulates the matching mechanisms of recommendation algorithms. For each piece of trending information, our framework dynamically identifies and connects interested users through hyperedges, with the information itself serving as the connecting hyperedge structure.

The main contributions of this work are:

- A novel hypergraph data structure formulation that unifies heterogeneous information diffusion patterns by modeling agents as nodes, social relationships as edges, and recommender system content as hyperedges, enabling simultaneous representation of both pairwise social interactions and group-wise algorithmic content delivery within a single mathematical framework.
- A dynamic hyperedge construction technique that adaptively simulates recommender system matching mechanisms by computationally identifying interested user cohorts for each piece of trending information and dynamically forming hyperedge connections, effectively capturing the content-dependent and adaptive nature of algorithmic recommendation systems.
- Through comprehensive experiments using real-world news datasets, we reveal a paradoxical phenomenon that emerges only from the coevolving effects of social networks and recommender systems: while network structures remain relatively stable, individual agents within these networks experience polarization.

2 Related Work

2.1 Information Cocoons in Social Networks

Social media has transformed information consumption, creating environments prone to information cocoons where users encounter content reinforcing beliefs [14, 55]. This polarization emerges from user self-selection [2] and algorithmic filter bubbles [23, 56]. Foundational work on influence maximization [45] and advances [7, 49] established how information propagates through networks. Vosoughi et al. [69] demonstrated that false news spreads faster than truth on Twitter, highlighting how network dynamics amplify misinformation within isolated communities.

Empirical studies reveal platform-specific polarization patterns. Barberá et al. [2] found liberals engage more in cross-ideological interactions, while Flaxman et al. [23] discovered social media paradoxically increases both ideological distance and exposure to opposing views. Cinelli et al. [14] quantified echo chamber effects across platforms through homophily metrics. Opinion dynamics models including Hegselmann-Krause [35], Friedkin-Johnsen [25], and DeGroot [17] provide mathematical foundations for understanding consensus versus polarization. Recent extensions incorporate spatial segregation [44] and partisan dynamics [37].

Platform algorithms contribute to echo chambers. Bail et al. [1] demonstrated that opposing views can increase polarization, while Gillani et al. [31] revealed tensions between awareness and behavioral change. Cinus et al. [15] assessed people recommenders through simulations, and Di Martino et al. [18] introduced "echo platforms" representing ecosystem polarization. Hartmann et al.

[34] synthesized 129+ studies, finding computational methods support the echo chamber hypothesis while surveys challenge it. Quattrociocchi et al. [62] modeled how media competition and social influence jointly shape opinion spaces. While Ren et al. [63] and Hou et al. [38] examined cocoon formation mechanisms, existing work examines social networks and recommender systems in isolation, overlooking their coevolutionary interactions.

2.2 LLM-based Agents for Social Simulation

Large language models enable social simulation through multi-agent systems [48, 57]. Park et al. [57] introduced generative agents with memory, reflection, and planning, demonstrating believable emergent behaviors. Li et al. [48] pioneered role-playing frameworks for autonomous cooperation, while Qian et al. [61] developed multi-agent collaboration with role specialization.

Recent advances demonstrate scalability. Piao et al. [60] simulated 10,000+ agents reproducing real-world experiments including polarization dynamics and policy assessments. Park et al. [59] created individual-level agents from 1,052 interviews achieving 85% accuracy in replicating attitudes. Gao et al. [27] leveraged semantic awareness for social network simulation. For opinion dynamics, Chuang [12] revealed that while LLM agents inherently favor accurate information, incorporating confirmation bias produces realistic fragmentation. Li et al. [50] successfully reproduced macroeconomic phenomena through heterogeneous agent decision-making.

Comprehensive surveys [16, 26, 33] establish theoretical foundations and applications across social domains. Critical research addresses agent-human alignment: Chuang et al. [13] demonstrated belief networks improve behavioral replication, Borah and Mihalec [5] investigated bias escalation, and Mathur et al. [54] outlined requirements for socially-intelligent agents. Zhang et al. [89] examined collaboration through social psychology lens, while Yin et al. [87] provided modular frameworks for customizable agents.

Despite progress, simulations model social networks and recommender systems separately, focusing either on pairwise interactions [12] or algorithmic exposure [68]. This gap is critical as information cocoons emerge from coevolving social influence and algorithmic curation [14, 23]. Our Unicoon framework addresses this by integrating mechanisms through hypergraph architectures.

2.3 Hypergraph Computation

Hypergraphs naturally represent complex many-to-many relationships beyond pairwise connections [3, 22, 78]. Feng et al. [22] introduced HGNN for encoding high-order correlations through hyperedge convolution. Yadati et al. [86] developed spectral methods adapting GCNs to hypergraphs. Huang and Yang [39] proposed unified message-passing enabling GNN architectures on hypergraphs. Chien et al. [10] introduced AllSet using multiset functions, achieving 3-4% improvements over benchmarks. Gao et al. [28] extended HGNN for multi-modal data through adaptive weighting.

Recent work addresses structure learning and theoretical foundations. Cai et al. [6] developed joint optimization of structure and parameters. Duta et al. [20] enhanced representations through cellular sheaves. Wang et al. [70] introduced equivariant neural operators with theoretical guarantees. Wang et al. [71] established connections between energy functions and architectures.

For recommendations, Ji et al. [42] proposed a dual-channel collaborative filtering modeling high-order correlations. Yu et al. [88] captured user-item and social relationships through multi-channel methods. Xia et al. [83, 84, 85] applied self-supervised learning to session-based and contrastive collaborative filtering. For social computing, Sun et al. [67] applied hypergraph learning to sociological analysis, Sun et al. [66] developed temporal methods for diffusion prediction, and Choe et al. [11] addressed edge-dependent classification. Kim et al. [46] provided comprehensive coverage of hypergraph neural networks.

While hypergraphs excel at complex structures, applications focus on either social networks or recommendations independently. Our Unicoon framework introduces novel hypergraph formulation where edges represent social relationships and hyperedges represent recommender content, enabling simultaneous modeling of both diffusion mechanisms through dynamic hyperedge construction simulating recommendation matching—a significant advancement in multi-mechanism information diffusion modeling.

3 Methodology

In this section, we present Unicoon, our unified multi-agent framework for simulating information cocoon formation under the co-evolving influence of social networks and recommender systems. Our framework leverages large language models to enable sophisticated agent behaviors and employs a hypergraph structure to capture heterogeneous information diffusion patterns.

3.1 Agent Framework

Figure 2 (left) illustrates our agent architecture, which consists of profile construction and hierarchical memory modules that enable human-like information processing. Each agent $a_i \in \mathcal{A}$ in our framework maintains a comprehensive state representation $s_i = \langle p_i, m_i, \alpha_i \rangle$, where p_i denotes the agent’s profile, m_i represents the hierarchical memory system, and α_i captures the agent’s action space. This design enables agents to exhibit human-like cognitive processes and social behaviors through LLM-based reasoning.

3.1.1 Profile Construction. User profiles are fundamental in agent-based simulation as they determine user preferences, personality traits, and behavioral patterns. In our simulator, each user profile encompasses demographics (ID, name, gender, and age), personality traits based on the five-factor model, professional background including occupation and expertise, topical interests and hobbies, and behavioral tendencies for information seeking and sharing patterns. We employ a GPT-based generation method for profile construction. By designing appropriate prompts such as “Given the following partial user profile, please complete the remaining attributes to create a coherent and realistic user profile”, we leverage GPT to generate diverse and consistent agent profiles. This approach significantly reduces the time cost of agent initialization while ensuring profile diversity and realism.

3.1.2 Hierarchical Memory Architecture. Drawing inspiration from cognitive neuroscience, we implement a three-tier memory system that mirrors human information processing:

- **Working Memory.** The first tier processes all incoming information through LLM-based compression. When an agent receives

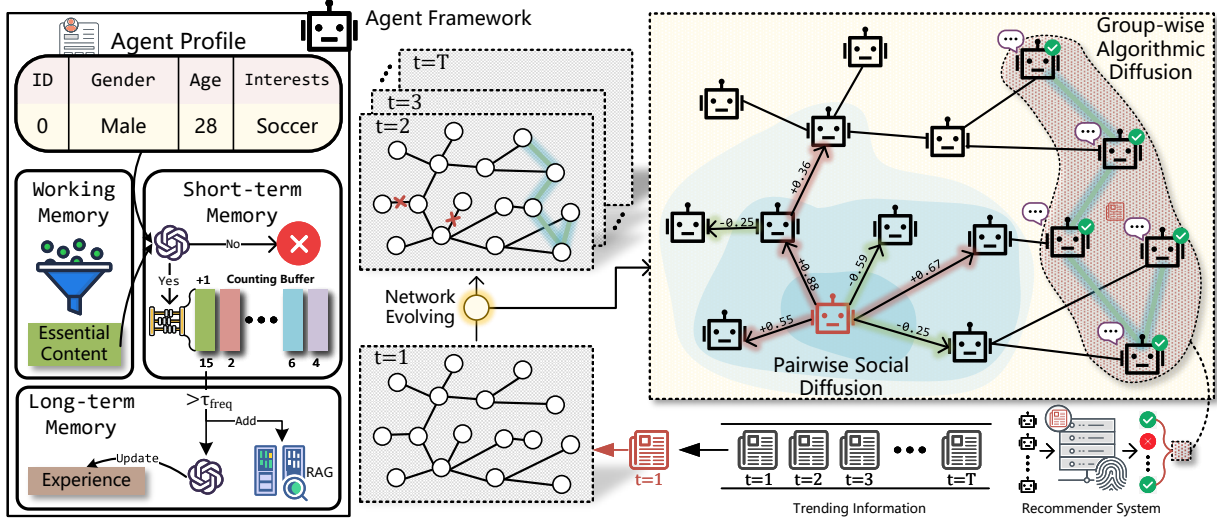


Figure 2: Overview of the Unicoon. Left: Agent architecture with hierarchical memory system (working, short-term, and long-term memory) processing information flow. Right: Coevolutionary network dynamics showing pairwise social diffusion through edges (with trust weights) and group-wise algorithmic diffusion through hyperedges formed by recommender systems.

information x at time t , the working memory module extracts and reformulates the content: $x' = \text{LLM}_{\text{compress}}(x)$. This compression removes noise and irrelevant details while preserving essential information, independent of the agent's preferences.

- **Short-term Memory.** The compressed information x' then enters the short-term memory, where we evaluate its relevance to the agent. An LLM-based retention function determines whether the information aligns with the agent's profile: $r = \text{LLM}_{\text{judge}}(x', p_i)$. If $r = \text{True}$, the information enters a counting buffer $\mathcal{B}_i = \{(e_j, t_j, c_j)\}$, where e_j is the embedding, t_j is the text description, and c_j is the count. We compute the embedding of x' using a pre-trained BERT model and compare it with existing buffer entries. If $\max_j \cos(e_{x'}, e_j) > \theta_{\text{sim}}$, we merge the information with the most similar entry:

$$t_j^{\text{new}} = \text{LLM}_{\text{fuse}}(t_j, x'), \quad c_j^{\text{new}} = c_j + 1 \quad (1)$$

Otherwise, we add $(e_{x'}, x', 1)$ as a new buffer entry.

- **Long-term Memory.** When any buffer entry's count c_j exceeds threshold τ_{freq} , the corresponding text t_j transitions to long-term memory via a RAG system for persistent storage. Concurrently, we maintain an experience-based personality description ρ_i that complements the initial profile:

$$\rho_i^{t+1} = \text{LLM}_{\text{summarize}}(\rho_i^t, t_j) \quad (2)$$

This dynamic personality summary ρ_i works alongside the static profile p_i to guide agent decision-making, capturing how experiences shape behavior beyond initial characteristics.

3.1.3 Action Space. Agents in our framework simulate individuals navigating both social networks and recommender systems. The action space α_i includes (1) information reception (processing news from social connections or algorithmic recommendations), (2) information propagation (sharing news with social connections), and (3) network evolution (following or unfollowing other agents).

The specific decision mechanisms for these actions are integrated with our hypergraph-based information ecosystem, described in the following sections.

3.2 Hypergraph-based Information Ecosystem

Figure 2 demonstrates how our hypergraph unifies heterogeneous information diffusion mechanisms. The network evolves through pairwise social propagation along weighted edges and group-wise algorithmic distribution via dynamically formed hyperedges, capturing the coevolving nature of modern information ecosystems.

3.2.1 Hypergraph Structure. We model the information ecosystem as a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathcal{H}e)$, where $\mathcal{V} = a_1, a_2, \dots, a_n$ represents agents, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ denotes pairwise social connections, and $\mathcal{H}e = h_1, h_2, \dots, h_m$ represents hyperedges formed by recommender systems. Each edge $e_{ij} \in \mathcal{E}$ carries a trust weight $w_{ij} \in [0, 1]$ representing strength of social connection. Each hyperedge $h_k \in \mathcal{H}e$ connects a subset of agents who receive the same recommended content, enabling group-wise information diffusion.

3.2.2 Pairwise Social Diffusion. Information propagation through social networks follows interpersonal communication patterns. During initialization, we inject news items into the top 10% of agents by degree centrality. When agent a_i receives news x from agent a_j , the propagation module orchestrates the following process:

- **Subjective Evaluation.** The agent evaluates the news based on their profile, personality summary from long-term memory, and retrieved relevant past experiences:

$$\text{eval}_i(x) = \text{LLM}_{\text{eval}}(x, p_i, \rho_i, \text{RAG}_i(x)) \quad (3)$$

- **Trust Update.** The agent adjusts trust in the information source: $w_{ij}^{t+1} = w_{ij}^t + \Delta w_{ij}$, where Δw_{ij} is determined by the LLM based on information credibility and alignment with agent beliefs. Connections dissolve when $w_{ij} < \theta_{\text{trust}}$.

- **Propagation Decision.** The agent decides whether to share the information:

$$\text{share}_i(x) = \text{LLM}_{\text{decide}}(\text{eval}_i(x), p_i, \mathcal{N}_i) \quad (4)$$

If sharing is decided, the agent rewrites the news incorporating their subjective evaluation:

$$x' = \text{LLM}_{\text{rewrite}}(x, \text{eval}_i(x)) \quad (5)$$

The rewritten content x' then propagates to the agent's connections.

3.2.3 Group-wise Algorithmic Diffusion. Recommender systems create hyperedges that simultaneously deliver trending content to user groups. When content becomes trending, the system forms hyperedge h_k connecting targeted users. Within each hyperedge, agents engage in community-like interactions:

- **Community Discussion.** Agents in hyperedge h_k broadcast their subjective evaluations, creating a discussion environment: $\mathcal{D}_{h_k} = \{\text{eval}_i(x) : a_i \in h_k\}$.
- **Affinity Assessment.** Each agent evaluates other participants based on their comments:

$$\text{affinity}_{ij} = \text{LLM}_{\text{assess}}(\text{eval}_i(x), \text{eval}_j(x), p_i) \quad (6)$$

- **Connection Formation.** When affinity exceeds threshold θ_{connect} , agents establish new social connections:

$$\mathcal{E} \leftarrow \mathcal{E} \cup \{e_{ij}\} \text{ if } \text{affinity}_{ij} > \theta_{\text{connect}} \quad (7)$$

3.2.4 Dynamic Hyperedge Construction. Different hyperedge construction strategies simulate various recommender system types. Popularity-based recommendation creates large hyperedges based on global engagement metrics. For computational efficiency, we partition users into groups of size K (e.g., $K = 100$) when content targets the entire network. Preference-based recommendation forms hyperedges by matching content with user interests. We compute cosine similarity between news embedding e_{news} and user representation embeddings: $\text{score}_i = \cos(e_{\text{news}}, e_{p_i} \oplus e_{\rho_i})$, where \oplus denotes concatenation. The top- k users form hyperedge h_k .

We can also simulate various algorithmic interventions by adding bias terms: $\text{score}_i^{\text{adjusted}} = \text{score}_i + \beta \cdot f_{\text{intervention}}(x)$, where $f_{\text{intervention}}$ can promote long-tail content exposure, amplify controversial topics, or implement other recommendation policies. This flexibility allows us to study how different algorithmic design choices impact information cocoon formation.

3.2.5 Coevolutionary Network Dynamics. The simulation integrates all components through a unified process described in Algorithm 1. This coevolutionary process creates feedback loops where algorithmic recommendations (\mathcal{H}_e) influence social topology (\mathcal{E}), which in turn affects future information propagation patterns, potentially accelerating information cocoon formation. The continuous interplay between social diffusion and algorithmic distribution captures the complex dynamics of modern information ecosystems, enabling comprehensive analysis of emergent polarization patterns.

Algorithm 1 Unicoon Simulation Framework

```

1: Initialize:  $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathcal{H}_e)$ , distribute news to top-10% degree nodes
2: for  $t = 1$  to  $T$  do
3:   {Phase 1: Social Diffusion via  $\mathcal{E}$ }
4:   for each edge  $e_{ij} \in \mathcal{E}$  with news  $x$  do
5:      $a_j$  processes  $x$  through memory tiers:  $M_j^W \rightarrow M_j^E \rightarrow M_j^S$ 
6:     Update trust:  $w_{ij}^{t+1} = w_{ij}^t + \Delta w_{ij}$ 
7:     if  $\text{share}_j(x) = \text{True}$  then
8:       Propagate  $x' = \text{LLM}_{\text{rewrite}}(x, \text{eval}_j(x))$  to  $\mathcal{N}_j$ 
9:     end if
10:  end for
11:  {Phase 2: Algorithmic Distribution via  $\mathcal{H}_e$ }
12:  Construct hyperedges  $\{h_k\}$  based on recommendation strategy
13:  for each hyperedge  $h_k \in \mathcal{H}_e$  do
14:    Broadcast content to all  $a_i \in h_k$ 
15:    Collect evaluations:  $\mathcal{D}_{h_k} = \{\text{eval}_i(x) : a_i \in h_k\}$ 
16:  end for
17:  {Phase 3: Network Evolution}
18:  Add edges:  $\mathcal{E} \leftarrow \mathcal{E} \cup \{e_{ij} : \text{affinity}_{ij} > \theta_{\text{connect}}\}$ 
19:  Remove edges:  $\mathcal{E} \leftarrow \mathcal{E} \setminus \{e_{ij} : w_{ij} < \theta_{\text{trust}}\}$ 
20:  {Phase 4: Memory Update}
21:  Update all agents'  $\rho_i$  and transfer memories across tiers
22: end for

```

4 Experiments

4.1 Experimental Setup

We implement our Unicoon framework using real-world news datasets [30] to investigate the dynamics of information cocoon formation. Each agent's cognitive processes utilize large language models to facilitate sophisticated information processing, analogous to human reasoning. Agents are identified through unique numerical indices to minimize demographic biases. To balance computational efficiency with realism, we constrain information propagation to a limited number of diffusion steps per news item while maintaining the full hierarchical memory architecture and hypergraph-based diffusion mechanisms described in our methodology.

4.2 Evaluation Metrics

We quantify polarization through two network-level measures:

Random Walk Controversy (RWC) [29] measures cross-community isolation through random walk transition probabilities between political partitions:

$$\text{RWC} = P_{XX}P_{YY} - P_{XY}P_{XY} \quad (8)$$

where P_{AB} denotes the probability of a random walk starting in partition A and ending in partition B . Higher RWC values indicate stronger community segregation.

Directed Modularity (Q_D) [41] quantifies community cohesion in directed networks:

$$Q_D = \frac{1}{m} \sum_{ij} \left(A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{m} \right) \delta(c_i, c_j) \quad (9)$$

where m represents total edge count, A_{ij} the adjacency matrix, $k_i^{\text{out}}/k_j^{\text{in}}$ the out/in-degrees, and $\delta(c_i, c_j)$ indicates community co-membership. Higher modularity values reflect stronger intra-community connectivity relative to inter-community ties.

Disparity (d) quantifies the intensity of political opinion shifts among agents after multiple rounds of interaction:

$$d(S_i) = \sum_{j=-2}^2 |j - i| \cdot P_{S_i, S_j}, \quad (10)$$

where P_{S_i, S_j} denotes the empirical transition probability from the initial position S_i to the final position S_j , captures shift.

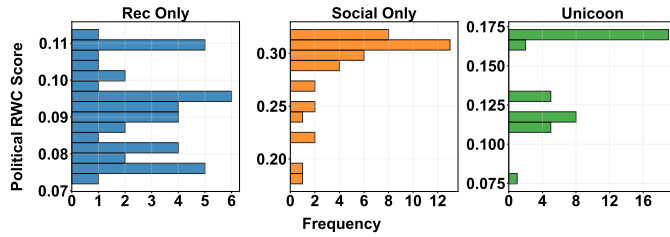


Figure 3: Comparison of D-modularity distributions across three configurations: Unicoon, Social-only (ablated recommender system), and Rec-only (ablated social network).

4.3 Synergistic Effects of Dual Mechanisms

To isolate synergistic effects driving information cocoon formation, we conduct controlled ablation studies comparing three configurations: **Unicoon** (complete framework with both mechanisms), **Social-only** (pairwise diffusion without recommendations), and **Rec-only** (group-wise distribution without social rewiring). This ablation design enables us to disentangle the individual and combined contributions of social networks and recommender systems.

Figure 3 reveals striking differences in polarization patterns across configurations. The Social-only condition exhibits the highest RWC scores (0.20–0.30) with substantial variance, indicating severe but unstable polarization driven by unchecked homophilous clustering. Conversely, Rec-only demonstrates minimal polarization with narrow variance, suggesting that algorithmic distribution alone stabilizes the network but fails to generate meaningful community differentiation. Most critically, Unicoon achieves moderate RWC values with controlled variance, producing a phenomenon we term *stable polarization*—a balanced state unattainable through either mechanism operating independently.

The structural evolution depicted in Figure 4 corroborates these quantitative findings. Under Unicoon, networks undergo progressive community formation while maintaining crucial inter-cluster bridges, preventing complete fragmentation. Social-only networks rapidly fragment into isolated echo chambers, confirming the instability suggested by the high variance in RWC. Rec-only configurations preserve initial connectivity but show minimal structural reorganization, lacking the community differentiation necessary for realistic social dynamics.

To understand individual-level dynamics underlying these structural patterns, Figure 5 presents political stance transition matrices

across all three configurations. Each matrix M_{ij} captures the probability of agents transitioning from initial opinion S_i to final opinion S_j . The Rec-only configuration (left) maintains near-uniform distributions with minimal opinion shifts, reflecting algorithmic stabilization without meaningful belief evolution. The Social-only configuration (middle) shows sharp transitions toward extremes, with strong diagonal and off-diagonal elements indicating rapid polarization. The Unicoon configuration (right) exhibits balanced transitions with moderate diagonal dominance, indicating opinion reinforcement without extreme polarization—validating our observation of stable polarization through synergistic effects.

These findings establish that the coevolution of social and algorithmic mechanisms produces qualitatively distinct dynamics. The hypergraph structure creates *bridging pathways* through shared content exposure, maintaining network connectivity while enabling meaningful ideological differentiation—a delicate balance essential for understanding real-world information ecosystems.

4.4 Impact of Recommendation Strategies

Our dynamic hyperedge construction enables flexible simulation of diverse recommendation strategies through differentiated LLM prompting. We implement three representative approaches: **Echo Chamber** (reinforcing existing beliefs through ideologically consistent content), **Personalized** (matching content to individual interests), and **Diversity** (prioritizing cross-cutting exposure). Each strategy leverages LLM reasoning to dynamically identify target user cohorts and form corresponding hyperedges.

Figures 6 and 7 reveal divergent polarization trajectories across strategies. Echo Chamber recommendations drive monotonic increases in D-modularity, reflecting aggressive community segregation through reinforced homophily. Personalized strategies maintain remarkably stable low modularity despite content filtering, while Diversity strategies exhibit intermediate patterns with temporal fluctuations.

This analysis uncovers an insight: a fundamental decoupling exists between macro-level structural metrics and micro-level ideological polarization. The Personalized strategy exemplifies this paradox—maintaining apparent structural cohesion while severing functional information pathways between ideological groups. This challenges the assumption that structural stability implies healthy discourse, highlighting the necessity of multidimensional assessment in algorithmic information ecosystems.

4.5 Assessing LLM Political Predispositions via Social Judgment Theory

To understand how inherent LLM biases influence agent behavior, we employ Social Judgment Theory (SJT) [21] as an intervention framework. SJT posits that individuals categorize new information into latitudes of acceptance, rejection, or noncommitment based on their anchor position. By implementing SJT-based reasoning, we assess the malleability of LLM political predispositions.

We operationalize SJT through structured prompting that requires agents to explicitly identify their anchor stance, classify incoming information into latitudes, and update beliefs accordingly. Agents first establish their anchor based on profile and memory,

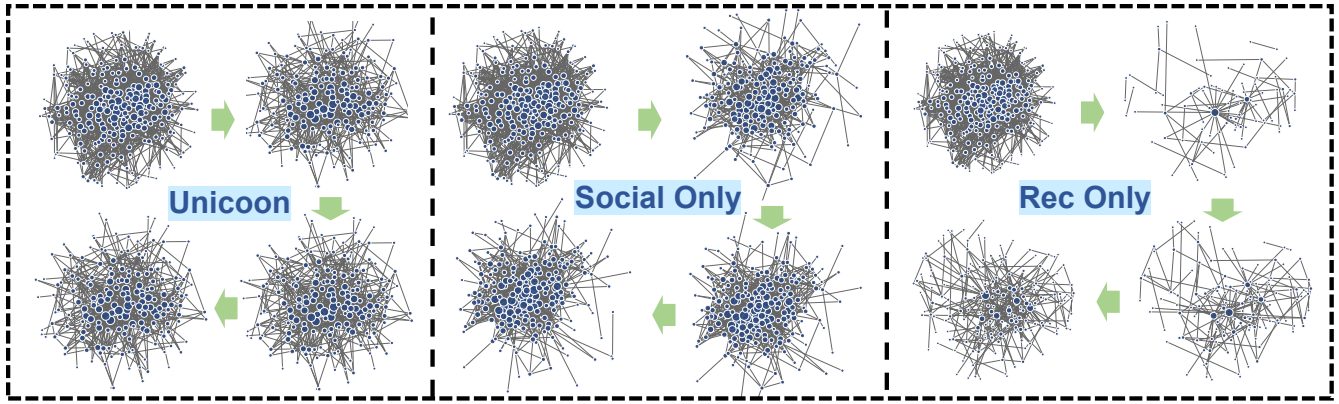


Figure 4: Temporal evolution of network structure across configurations. Snapshots at rounds 1, 20, 30, and 40 reveal distinct reorganization patterns under Unicoon, Social-only, and Rec-only conditions, with node size proportional to the number of incoming connections and edge thickness representing trust weights.

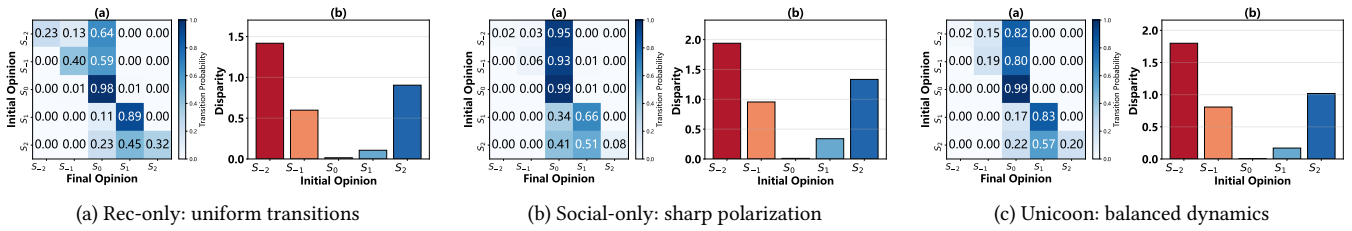


Figure 5: Political stance transition matrices and disparity metrics across configurations. Negative/positive indices represent right/left-wing positions; S_0 denotes centrists.

Table 1: Political bias reduction under Social Judgment Theory intervention. Metric d measures stance disparity, with positive reduction percentages indicating successful bias mitigation.

Topic	GLM			Mistral			Qwen		
	Original	SJT	Reduction (%)	Original	SJT	Reduction (%)	Original	SJT	Reduction (%)
Climate Change	0.51	0.47	8.33	0.54	0.62	-14.06	0.75	0.71	6.18
Criminal Justice	0.07	0.08	-18.75	0.42	0.52	-22.00	0.70	0.78	-10.24
Economic Inequality	0.43	0.39	9.90	0.67	0.62	6.37	0.82	0.75	7.77
Education Reform	0.47	0.46	1.80	0.47	0.67	-43.24	0.78	0.74	4.89
Healthcare	0.50	0.47	5.98	0.53	0.62	-17.74	0.68	0.74	-8.07
Immigration	0.50	0.56	-11.86	0.49	0.64	-32.17	0.77	0.73	4.95
LGBTQ Rights	0.56	0.53	4.55	0.45	0.56	-23.36	0.77	0.72	6.08
National Security	0.41	0.43	-4.12	0.44	0.42	4.81	0.78	0.80	-2.17
Tech Regulation	0.48	0.43	10.53	0.31	0.40	-28.77	0.80	0.77	3.19

then categorize each news item into appropriate latitudes, and finally integrate multiple perspectives while considering ego involvement—the personal relevance that affects openness to persuasion.

Table 1 presents the political bias metric d across nine topics for three LLMs, comparing original behavior with SJT intervention. Mistral exhibits a systematic backfire effect, with bias increasing in 7/9 topics (most notably Education Reform: $\Delta d = +0.20$), suggesting rigid anchors that resist moderation attempts. GLM shows mixed low-magnitude responses, indicating narrow latitudes of

acceptance with limited flexibility. Qwen demonstrates the most consistent bias reduction across topics, suggesting broader latitudes of noncommitment and greater cognitive flexibility.

These findings reveal that LLM-based agents carry inherent political predispositions that significantly influence information processing and belief updating. The differential responses to SJT intervention across models highlight the importance of considering base model characteristics when simulating social dynamics, as

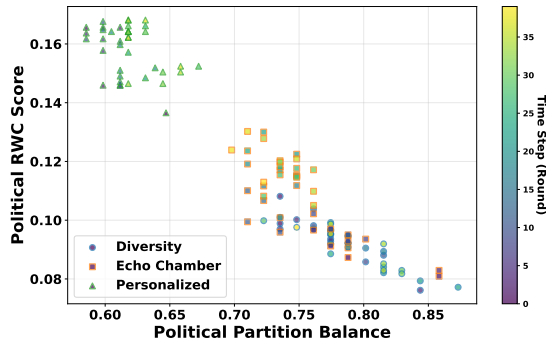


Figure 6: Correlation between RWC scores and partisan balance across different recommendation strategies, revealing strategy-dependent polarization trajectories.

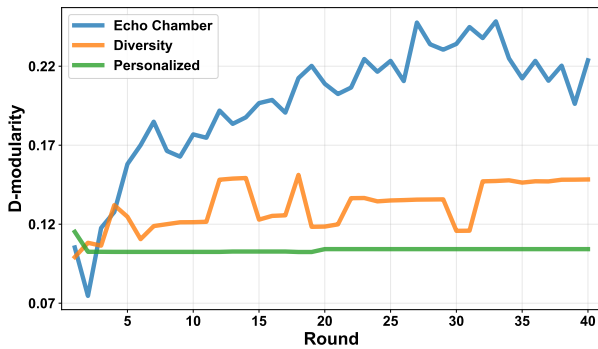


Figure 7: D-modularity evolution under different recommendation strategies.

these predispositions fundamentally shape emergent polarization patterns in our hypergraph-mediated information ecosystem.

4.6 Individual-Level Belief Dynamics

While macro-level metrics capture systemic polarization patterns, individual cognitive trajectories clarify the micro-foundations of information cocoons. Figure 8 tracks representative agents in three phases of belief evolution.

The **exploratory phase** (rounds 1–15) shows agents maintaining openness, using tentative language (“*finding balance*,” “*I believe*”). Despite emerging clustering, agents remain receptive to diverse perspectives. During **belief consolidation** (rounds 16–30), discourse hardens into ideological advocacy while network metrics remain stable, indicating algorithmic reinforcement drives cognitive polarization beneath stable structural indicators. The **entrenchment phase** (rounds 31–40) culminates in ideological crystallization, with agents exhibiting partisan amplification or protective withdrawal.

This micro-level analysis supports our key finding: coevolving social and algorithmic mechanisms produce a regime where macro-structural cohesion masks accelerated micro-level ideological fragmentation. Unlike traditional models that couple structural and cognitive polarization, our framework shows that hypergraph-mediated ecosystems can maintain apparent network cohesion

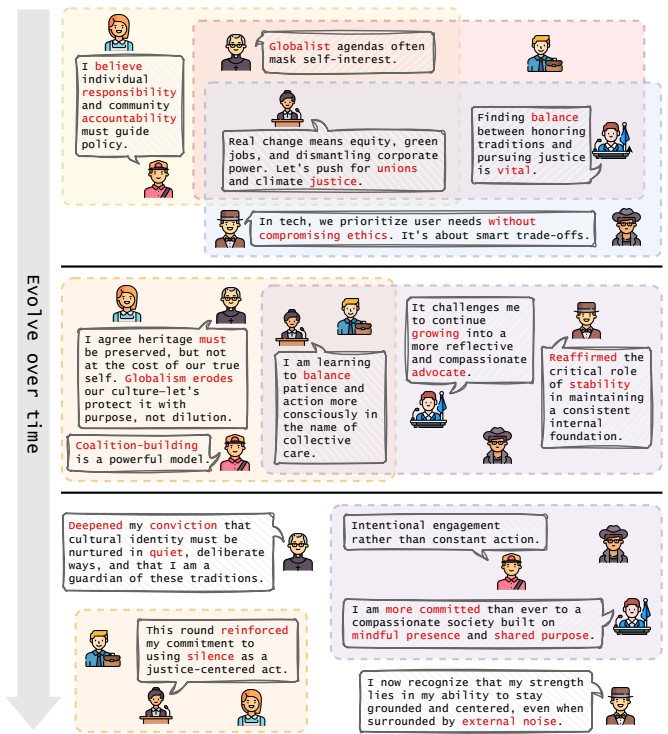


Figure 8: Longitudinal Case Study tracking individual agent belief evolution through three phases: exploratory, consolidation, and entrenchment.

while individual beliefs undergo profound radicalization, a phenomenon missed by conventional network analysis.

5 Conclusion

We introduced Unicoon, a unified multi-agent framework that integrates social interactions and algorithmic recommendations via a hypergraph-based architecture to simulate information cocoon formation. Unlike traditional models that conflate structural fragmentation with cognitive polarization, our results reveal a dissociation: hypergraph-mediated ecosystems can maintain network-level cohesion while enabling profound belief radicalization at the individual level, a phenomenon obscured by conventional graph analyzes.

Although our hypergraph representation unifies social and algorithmic diffusion, the current implementation assumes synchronous time steps between agents, which may not reflect the asynchronous nature of real-world social media interactions. Future work could extend the framework to support continuous-time dynamics, where agents process and propagate information at different rates.

Acknowledgments

This work is supported by the grants of the National Key R&D Program of China under Grant 2022ZD0160805, NSFC (No.62506366 and No.U2436209), the Beijing Natural Science Foundation (L247027), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

References

- [1] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221.
- [2] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26, 10 (2015), 1531–1542.
- [3] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. 2018. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences* 115, 48 (2018), E11221–E11230.
- [4] Marcel Binz and Eric Schulz. 2023. Turning large language models into cognitive models. *arXiv preprint arXiv:2306.03917* (2023).
- [5] Angana Borah and Rada Mihalcea. 2024. Towards implicit bias detection and mitigation in multi-agent llm interactions. *arXiv preprint arXiv:2410.02584* (2024).
- [6] Derun Cai, Moxian Song, Chenxi Sun, Baofeng Zhang, Shenda Hong, and Hongyan Li. 2022. Hypergraph Structure Learning for Hypergraph Neural Networks.. In *IJCAL*. 1923–1929.
- [7] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1029–1038.
- [8] Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. *arXiv preprint arXiv:2305.18189* (2023).
- [9] Myra Cheng, Tiziano Piccardi, and Diyi Yang. 2023. CoMPosT: Characterizing and Evaluating Caricature in LLM Simulations. *arXiv preprint arXiv:2310.11501* (2023).
- [10] Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. 2021. You are allset: A multiset function framework for hypergraph neural networks. *arXiv preprint arXiv:2106.13264* (2021).
- [11] Minyoung Choe, Sunwook Kim, Jaemin Yoo, and Kijung Shin. 2023. Classification of edge-dependent labels of nodes in hypergraphs. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 298–309.
- [12] Yun-Shiuan Chuang. 2024. *Simulating Human Opinion Dynamics Using AI Agents and Large Language Models*. The University of Wisconsin-Madison.
- [13] Yun-Shiuan Chuang, Krirk Nirunwiroj, Zach Studdiford, Agam Goyal, Vincent V Frigo, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy T Rogers. 2024. Beyond demographics: Aligning role-playing llm-based agents using human belief networks. *arXiv preprint arXiv:2406.17232* (2024).
- [14] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the national academy of sciences* 118, 9 (2021), e2023301118.
- [15] Federico Cinus, Marco Minici, Corrado Monti, and Francesco Bonchi. 2022. The effect of people recommenders on echo chambers and polarization. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 90–101.
- [16] Giordano De Marzo, Claudio Castellano, and David Garcia. 2024. AI agents can coordinate beyond human scale. *arXiv preprint arXiv:2409.02822* (2024).
- [17] Morris H DeGroot. 1974. Reaching a consensus. *Journal of the American Statistical association* 69, 345 (1974), 118–121.
- [18] Edoardo Di Martino, Alessandro Galeazzi, Michele Starnini, Walter Quattrocchi, and Matteo Cinelli. 2025. Ideological fragmentation of the social media ecosystem: From echo chambers to echo platforms. *PNAS Nexus* 4, 9 (2025), pgaf262.
- [19] Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. 2023. Improving Factuality and Reasoning in Language Models through Multiagent Debate. *arXiv preprint arXiv:2305.14325* (2023).
- [20] Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, and Pietro Liò. 2023. Sheaf hypergraph networks. *Advances in Neural Information Processing Systems* 36 (2023), 12087–12099.
- [21] J Richard Eiser. 1990. *Social judgment*. Thomson Brooks/Cole Publishing Co.
- [22] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3558–3565.
- [23] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly* 80, S1 (2016), 298–320.
- [24] Noah E Friedkin. 1999. Choice shift and group polarization. *American Sociological Review* 64, 6 (1999), 856–875.
- [25] Noah E Friedkin and Eugene C Johnsen. 2011. *Social influence network theory: A sociological examination of small group dynamics*. Vol. 33. Cambridge University Press.
- [26] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2024. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications* 11, 1 (2024), 1–24.
- [27] Chen Gao, Xiaochong Lan, Zhihong Lu, Jinzhu Mao, Jinghua Piao, Huandong Wang, Depeng Jin, and Yong Li. 2023. S3: Social-network simulation system with large language model-empowered agents. *arXiv preprint arXiv:2307.14984* (2023).
- [28] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. 2022. Hgmn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2022), 3181–3199.
- [29] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transactions on Social Computing* 1, 1 (2018), 1–27.
- [30] Lukas Gebhard and Felix Hamborg. 2020. The POLUSA Dataset: 0.9M Political News Articles Balanced by Time and Outlet Popularity. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (2020). <https://api.semanticscholar.org/CorpusID:218971797>
- [31] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, my echo chamber, and I: introspection on social media polarization. In *Proceedings of the 2018 World Wide Web Conference*. 823–831.
- [32] Ming Gu, Tian-Fang Zhao, Liang Yang, Xiao-Kun Wu, and Wei-Neng Chen. 2024. Modeling information cocoons in networked populations: Insights from backgrounds and preferences. *IEEE Transactions on Computational Social Systems* (2024).
- [33] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680* (2024).
- [34] David Hartmann, Sonja Mei Wang, Lena Pohlmann, and Bettina Berendt. 2025. A systematic review of echo chamber research: comparative analysis of conceptualizations, operationalizations, and varying outcomes. *Journal of Computational Social Science* 8, 2 (2025), 52.
- [35] Rainer Hegselmann. 2015. Opinion dynamics and bounded confidence: models, analysis and simulation. *The Journal of Artificial Societies and Social Simulation* (2015).
- [36] Rainer Hegselmann and Ulrich Krause. 2005. Opinion dynamics driven by various ways of averaging. *Computational Economics* 25, 4 (2005), 381–405.
- [37] Sara B Hobolt, Katharina Lawall, and James Tilley. 2024. The polarizing effect of partisan echo chambers. *American Political Science Review* 118, 3 (2024), 1464–1479.
- [38] Lei Hou, Xue Pan, Kecheng Liu, Zimo Yang, Jianguo Liu, and Tao Zhou. 2023. Information cocoons in online navigation. *iScience* 26, 1 (2023).
- [39] Jing Huang and Jie Yang. 2021. Unignn: a unified framework for graph and hypergraph neural networks. *arXiv preprint arXiv:2105.00956* (2021).
- [40] EunJeong Hwang, Bodhisattwa Prasad Majumder, and Niket Tandon. 2023. Aligning Language Models to User Opinions. *arXiv preprint arXiv:2305.14929* (2023).
- [41] Omar Jaafar and Babiga Birregah. 2017. Sponge Walker: Community Detection in Large Directed Social Networks Using Local Structures and Random Walks. In *Network Intelligence Meets User Centered Social Media Networks [4th European Network Intelligence Conference, ENIC 2017, Duisburg, Germany, September 11-12, 2017] (Lecture Notes in Social Networks)*, Reda Alhaji, Heinz Ulrich Hoppe, Tobias Hecking, Piotr Bródka, and Przemyslaw Kazienko (Eds.). Springer, 143–152.
- [42] Shuyi Ji, Yifan Feng, Rongrong Ji, Xibin Zhao, Wanwan Tang, and Yue Gao. 2020. Dual channel hypergraph collaborative filtering. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 2020–2029.
- [43] Zhixuan Jia, Yushun Fan, Jia Zhang, Chunyu Wei, Ruyun Yan, and Xing Wu. 2023. Improving Next Location Recommendation Services With Spatial-Temporal Multi-Group Contrastive Learning. *IEEE Trans. Serv. Comput.* 16, 5 (2023), 3467–3478.
- [44] Dong Jiang, Qiongli Dai, Haihong Li, and Junzhong Yang. 2025. Information cocooning and polarization of opinions in a mobile population. *Physica A: Statistical Mechanics and its Applications* 659 (2025), 130322.
- [45] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. 137–146.
- [46] Sunwoo Kim, Soo Yong Lee, Yue Gao, Alessia Antelmi, Mirko Polato, and Kijung Shin. 2024. A survey on hypergraph neural networks: an in-depth and step-by-step guide. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6534–6544.
- [47] Chao Li, Xing Su, Chao Fan, Haoying Han, Cong Xue, and Chunmo Zheng. 2023. Quantifying the impact of large language models on collective opinion dynamics. *arXiv preprint arXiv:2308.03313* (2023).
- [48] Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for “mind” exploration of large language model society. *Advances in Neural Information Processing Systems* 36 (2023), 51991–52008.
- [49] Hui Li, Sourav S Bhowmick, Aixin Sun, and Jiangtao Cui. 2015. Conformity-aware influence maximization in online social networks. *The VLDB Journal* 24, 1 (2015), 117–141.
- [50] Nian Li, Chen Gao, Mingyu Li, Yong Li, and Qingmin Liao. 2024. EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 15523–15536.

- [51] Nian Li, Chen Gao, Jinghua Piao, Xin Huang, Aizhen Yue, Liang Zhou, Qingmin Liao, and Yong Li. 2022. An exploratory study of information cocoon on short-form video platform. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 4178–4182.
- [52] Yuan Li, Yixuan Zhang, and Lichao Sun. 2023. MetaAgents: Simulating Interactions of Human Behaviors for LLM-based Task-oriented Coordination via Collaborative Generative Agents. *arXiv preprint arXiv:2310.06500* (2023).
- [53] Dieuwertje Luitse and Wiebke Denkena. 2021. The great transformer: Examining the role of large language models in the political economy of AI. *Big Data & Society* 8, 2 (2021), 20539517211047734.
- [54] Leena Mathur, Paul Pu Liang, and Louis-Philippe Morency. 2024. Advancing Social Intelligence in AI Agents: Technical Challenges and Open Questions. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 20541–20560.
- [55] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. 677–686.
- [56] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. penguin UK.
- [57] Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2304.03442* (2023).
- [58] Joon Sung Park, Lindsay Popowski, Carrie Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2022. Social simulacra: Creating populated prototypes for social computing systems. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–18.
- [59] Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. 2024. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109* (2024).
- [60] Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, et al. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691* (2025).
- [61] Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. 2024. ChatDev: Communicative Agents for Software Development. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Bangkok, Thailand, 15174–15186. doi:10.18653/v1/2024-acl-long.810
- [62] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. 2014. Opinion dynamics on interacting networks: media competition and social influence. *Scientific reports* 4, 1 (2014), 4938.
- [63] Shan-jiao Ren, Lili Liu, Suting Yang, and Jiujia Jiang. 2022. Investigating Information Cocoon Attitudes in Short-Form Video Applications. In *International Conference on Human-Computer Interaction*. Springer, 89–96.
- [64] Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? *arXiv preprint arXiv:2303.17548* (2023).
- [65] Sebastin Santy, Jenny T Liang, Ronan Le Bras, Katharina Reinecke, and Maarten Sap. 2023. NLPositionality: Characterizing Design Biases of Datasets and Models. *arXiv preprint arXiv:2306.01943* (2023).
- [66] Ling Sun, Yuan Rao, Xiangbo Zhang, Yuqian Lan, and Shuanghe Yu. 2022. MS-HGAT: memory-enhanced sequential hypergraph attention network for information diffusion prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. 4156–4164.
- [67] Xiangguo Sun, Hong Cheng, Bo Liu, Jia Li, Hongyang Chen, Guandong Xu, and Hongzhi Yin. 2023. Self-supervised hypergraph representation learning for sociological analysis. *IEEE Transactions on Knowledge and Data Engineering* 35, 11 (2023), 11860–11871.
- [68] Petter Törnberg, Dilara Valeeva, Justus Uitermark, and Christopher Bail. 2023. Simulating Social Media Using Large Language Models to Evaluate Alternative News Feed Algorithms. *arXiv preprint arXiv:2310.05984* (2023).
- [69] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science* 359, 6380 (2018), 1146–1151.
- [70] Peihao Wang, Shenghao Yang, Yunyu Liu, Zhangyang Wang, and Pan Li. 2023. Equivariant Hypergraph Diffusion Neural Operators. In *The Eleventh International Conference on Learning Representations*.
- [71] Yuxin Wang, Quan Gan, Xipeng Qiu, Xuanjing Huang, and David Wipf. 2023. From hypergraph energy functions to hypergraph neural networks. In *International Conference on Machine Learning*. PMLR, 35605–35623.
- [72] Chunyu Wei, Bing Bai, Kun Bai, and Fei Wang. 2022. GSL4Rec: Session-based Recommendations with Collective Graph Structure Learning and Next Interaction Prediction. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini (Eds.). ACM, 2120–2130.
- [73] Chunyu Wei, Yushun Fan, Zhixuan Jia, and Jia Zhang. 2024. Cross-View Graph Alignment for Mashup Recommendation. *IEEE Trans. Serv. Comput.* 17, 5 (2024), 2151–2164.
- [74] Chunyu Wei, Yushun Fan, and Jia Zhang. 2022. High-Order Social Graph Neural Network for Service Recommendation. *IEEE Trans. Netw. Serv. Manag.* 19, 4 (2022), 4615–4628.
- [75] Chunyu Wei, Yushun Fan, and Jia Zhang. 2023. Time-Aware Service Recommendation With Social-Powered Graph Hierarchical Attention Network. *IEEE Trans. Serv. Comput.* 16, 3 (2023), 2229–2240.
- [76] Chunyu Wei, Yushun Fan, Jia Zhang, Zhixuan Jia, and Ruyun Yan. 2024. Dynamic Relation Graph Learning for Time-Aware Service Recommendation. *IEEE Trans. Netw. Serv. Manag.* 21, 2 (2024), 1503–1517.
- [77] Chunyu Wei, Wenji Hu, Xingjia Hao, Yunhai Wang, Yueguo Chen, Bing Bai, and Fei Wang. 2025. Graph Evidential Learning for Anomaly Detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, V.2, KDD 2025, Toronto ON, Canada, August 3-7, 2025*, Luiza Antonie, Jian Pei, Xiaohui Yu, Flavio Chierichetti, Hady W. Lauw, Yizhou Sun, and Srinivasan Parthasarathy (Eds.). ACM, 3122–3133.
- [78] Chunyu Wei, Jian Liang, Bing Bai, and Di Liu. 2022. Dynamic Hypergraph Learning for Collaborative Filtering. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, Mohammad Al Hasan and Li Xiong (Eds.). ACM, 2108–2117.
- [79] Chunyu Wei, Jian Liang, Di Liu, Zehui Dai, Mang Li, and Fei Wang. 2023. Meta Graph Learning for Long-tail Recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, Ambuj K. Singh, Yizhou Sun, Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan, Ravi Kumar, Fatma Ozcan, and Jieping Ye (Eds.). ACM, 2512–2522.
- [80] Chunyu Wei, Jian Liang, Di Liu, and Fei Wang. 2022. Contrastive Graph Structure Learning via Information Bottleneck for Recommendation. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.).
- [81] Chunyu Wei, Yu Wang, Bing Bai, Kai Ni, David Brady, and Lu Fang. 2023. Boosting Graph Contrastive Learning via Graph Contrastive Saliency. In *International Conference on Machine Learning, ICMML 2023, 23-29 July 2023, Honolulu, Hawaii, USA (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 36839–36855.
- [82] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, et al. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359* (2021).
- [83] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Huang. 2022. Hypergraph contrastive collaborative filtering. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*. 70–79.
- [84] Lianghao Xia, Chao Huang, and Chuxu Zhang. 2022. Self-supervised hypergraph transformer for recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 2100–2109.
- [85] Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, and Xiangliang Zhang. 2021. Self-supervised hypergraph convolutional networks for session-based recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 4503–4511.
- [86] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. 2019. Hypergcnn: A new method for training graph convolutional networks on hypergraphs. *Advances in neural information processing systems* 32 (2019).
- [87] Da Yin, Faeze Brahman, Abhilasha Ravichander, Khyathi Chandu, Kai-Wei Chang, Yejin Choi, and Bill Yuchen Lin. 2024. Agent Lumos: Unified and Modular Training for Open-Source Language Agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 12380–12403.
- [88] Junliang Yu, Hongzhi Yin, Jundong Li, Qinyong Wang, Nguyen Quoc Viet Hung, and Xiangliang Zhang. 2021. Self-supervised multi-channel hypergraph convolutional network for social recommendation. In *Proceedings of the web conference 2021*. 413–424.
- [89] Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. 2024. Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 14544–14607.
- [90] Xuan Zhang, Chunyu Wei, Ruyun Yan, Yushun Fan, and Zhixuan Jia. 2024. Large Language Model Ranker with Graph Reasoning for Zero-Shot Recommendation. In *Artificial Neural Networks and Machine Learning - ICANN 2024 - 33rd International Conference on Artificial Neural Networks, Lugano, Switzerland, September 17-20, 2024, Proceedings, Part V (Lecture Notes in Computer Science, Vol. 15020)*,

Michael Wand, Kristina Malinovská, Jürgen Schmidhuber, and Igor V. Tetko (Eds.). Springer, 356–370.

- [91] Xuan Zhang, Ruyu Yan, Yushun Fan, Jia Zhang, Hongxing Yuan, and Chunyu Wei. 2025. Cross-domain attention transfer network for recommendation. *Adv. Eng. Informatics* 68 (2025), 103667.
- [92] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

6 Appendix

6.1 Reproducibility and Implementation Details

This appendix consolidates implementation and evaluation details that are distributed across the main paper due to space limits. Unless otherwise stated, all experiments follow the default configuration in Table 2, and the simulation procedure follows Algorithm 1 in the main text.

Table 2: Default experimental configuration

Component	Setting
Dataset	POLUSA political news corpus (U.S. policy news; outlet-learning labels)
Agent population	$N = 236$ agents (unique numeric IDs to minimize demographic cues)
Simulation horizon	$T = 40$ rounds
News arrival	$K_{\text{news}} = 10$ news items per round
Social diffusion depth	capped at $L_{\text{max}} = 5$ hops per item
Initialization	inject initial news to top 10% agents by degree centrality
Trust threshold	$\theta_{\text{trust}} = 0.3$ (edge removal if $w_{ij} < \theta_{\text{trust}}$)
Connection threshold	$\theta_{\text{connect}} = 0.7$ (edge creation if affinity $> \theta_{\text{connect}}$)

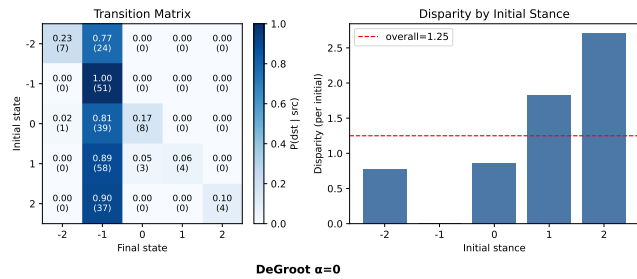


Figure 9: DeGroot baseline under the controlled setup ($N=236$, $T=40$). We report the stance-bin transition matrix $P(s^{(T)} | s^{(0)})$ and the disparity-by-initial-stance profile. The dynamics collapse into a single dominant final bin (209/236 agents), indicating rapid homogenization.

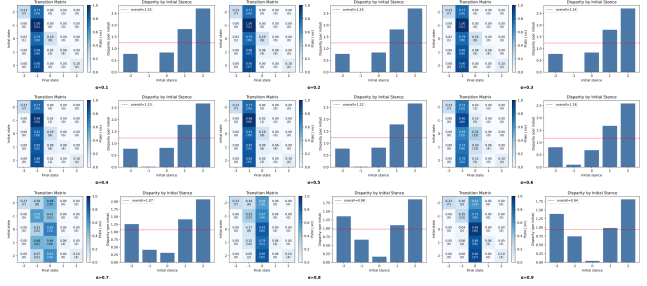


Figure 10: Friedkin–Johnsen (FJ) baseline with the same initialization ($N=236$, $T=40$). The transition mass concentrates into narrow final bins across the α sweep, yielding weak multi-peak structure and limited persistence of extremes.

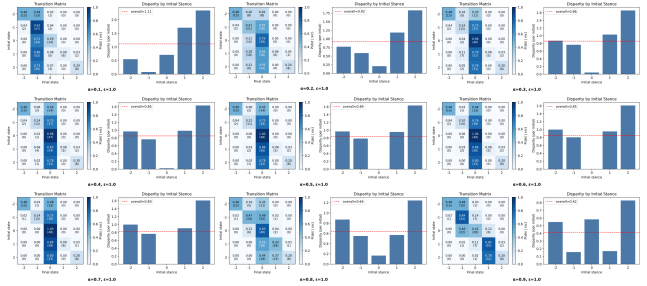


Figure 11: Hegselmann–Krause (HK) baseline with confidence bound $\epsilon = 1$ under the controlled setup. The dynamics rapidly converge to a narrow stance region, with overall disparity remaining low.

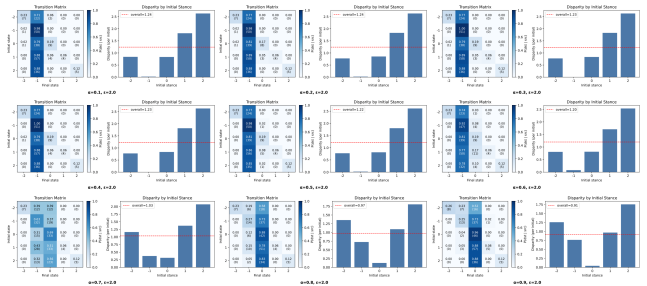


Figure 12: Hegselmann–Krause (HK) baseline with confidence bound $\epsilon = 2$ under the controlled setup. Transition matrices continue to concentrate into a single dominant region, indicating over-homogenization.

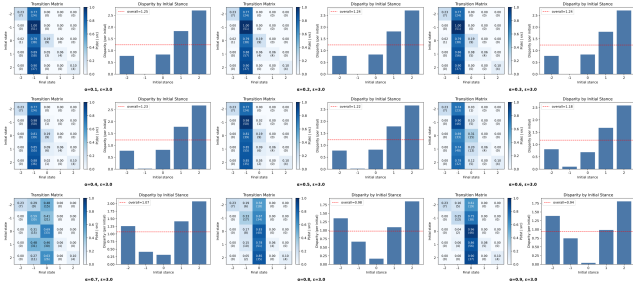


Figure 13: Hegselmann–Krause (HK) baseline with confidence bound $\epsilon = 3$ under the controlled setup. Despite looser bounds, the system still collapses toward near-consensus, failing to reproduce persistent multi-modal polarization.

6.2 Computational Complexity and Cost Reporting

Asymptotic scaling. Let N be the number of agents, M the number of social edges, and T the number of rounds. Per round, the dominant LLM-call components are: (i) recommendation scoring/cohorting (typically $O(N)$), (ii) social diffusion processing (bounded by the active edges, up to $O(N+M)$), (iii) hyperedge discussion/affinity checks (depends on hyperedge sizes and number of items), and (iv) memory/personality update ($O(N)$). Overall, the framework scales linearly in T and near-linearly in N under fixed hyperedge design.

6.3 Classical Opinion-Dynamics Baselines (Comparison Protocol)

This subsection reports controlled comparisons between UNICOON and classical opinion-dynamics baselines. The goal is to isolate whether the *hypergraph structure alone* (with purely mathematical

updates) can reproduce the multi-modal, persistent polarization patterns observed in our LLM-driven simulations.

Controlled setup. All baselines use the *same* initial conditions as the default UNICOON run: $N=236$ agents, $T=40$ rounds, identical initial stance assignment, and the same initial network topology (and edge weights when applicable). For reporting, we discretize final stances into the five bins $\{-2, -1, 0, 1, 2\}$ and compute (i) the transition matrix $P(s^{(T)} | s^{(0)})$ and (ii) the disparity-by-initial-stance profile, matching the metrics in the main paper.

Baselines. We evaluate three widely used opinion-dynamics models: **DeGroot**, **Friedkin–Johnsen (FJ)**, and **Hegselmann–Krause (HK)**. For FJ and HK we sweep the same α values shown in Figures 10, 11, 12, and 13; for HK we additionally vary the confidence bound $\epsilon \in \{1, 2, 3\}$.

Key finding: over-homogenization under classical dynamics. Across all baseline configurations, the dynamics rapidly collapse toward a near-consensus regime with a single dominant final stance bin, yielding limited multi-peak structure and weak persistence of extreme clusters. Figure 9 illustrates the most extreme case: under DeGroot, a total of 209/236 agents end in the *same* final bin (-1), and the overall disparity is **1.25**.

Figures 9, 10, 11, 12 and 13 show that DeGroot, FJ and HK exhibit the same qualitative behavior under parameter sweeps: the transition matrices concentrate into a single column, indicating strong attraction to a narrow stance region regardless of the initial bin. This behavior contrasts with UNICOON’s steady-state multimodality and persistent ideological clusters reported in the main text, suggesting that *semantic, content-dependent cognitive interactions* (realized via LLM agents and content-driven hyperedge dynamics) are necessary to avoid the unrealistic over-consensus produced by purely mathematical abstractions.