Simulating Social Networks with Large Language Models: The SNSim Framework

Anonymous ACL submission

Abstract

While large language models (LLMs) provide a promising technical foundation for social network simulation, existing approaches often implicitly equate simulation with text generation, thereby overlooking the dynamic interplay between opinion evolution and network structure. To address this limitation, we propose **SNSim**, a modular framework that formalizes social network simulation as the **joint and co-evolving** modeling of **language**, **opinion**, and **network structure**. SNSim adopts an **individual-community dual-channel** prompt design to systematically integrate personal attributes and community context into LLM-driven agents. Experiments on three real-world datasets demonstrate that SNSim consistently outperforms baseline approaches, successfully reproducing complex network topologies and temporally evolving opinion trajectories. Further analysis shows that the proposed modular design effectively mitigates linguistic homogenization while preserving structural realism. Taken together, these results indicate that SNSim provides a practical and operational framework to simulate social networks.

1 Introduction

As social interaction increasingly takes place on online platforms, social media has become a primary space where opinions form and spread. Simulating social networks therefore serves as a key approach in computational social science for studying opinion dynamics, polarization, and the impact of interventions under controlled conditions. Traditional simulation approaches typically rely on rule-based or equation-driven agent models, in which user behavior, opinion updates, and network evolution are governed by predefined heuristics. While these models offer interpretability, they often struggle to capture the richness of human language, context-sensitive, and identity-dependent nature of online interactions.

Recent advances in LLMs have renewed interest in social network simulation. With their ability to generate fluent and context-aware text, LLMs appear well suited to simulate social media users—producing posts, reacting to content, and engaging in discussions in ways that resemble human behavior (Park et al., 2022; Argyle et al., 2023). Consequently, a growing body of work has begun to employ LLMs as social agents (Zhou et al., 2024; Chuang et al., 2024), exploring tasks such as conversation generation, polarization dynamics, and synthetic social data creation. However, despite their promise, many existing LLM-based simulations remain ad hoc, with limited conceptual clarity regarding what is being simulated and how these components interact over time. Moreover, the use of LLMs for social simulation is still largely exploratory: commonly adopted components such as profiles or memory are often introduced in a heuristic manner, without a clear mapping between these mechanisms and the specific social or behavioral dimensions they are intended to model.

To move beyond these limitations, we clarify what is required to simulate a social network. A social network is not merely a collection of texts or interactions; rather, it emerges from the *joint interplay of three core components*: (1) **Language**, through which users express themselves via diverse lexical choices, semantic framing, and syntactic patterns; (2) **Opinion**, representing users' stances and beliefs toward certain topics; and (3) **Network structure**, which governs who communicates with whom and how social ties evolve. We hypothesize that by modeling these components and their interactions simultaneously we can better capture the dynamics observed in real-world social networks.

Building on this formulation, we introduce **SNSim**, a modular framework for LLM-based social network simulation. SNSim decomposes social simulation into three interdependent modules **Profile**, **Cognition**, and **Rewire**—corresponding to

language, opinion, and network structure, respectively. Across the three modules, SNSim incorporates both agents' individual perspectives and their surrounding community context into LLM prompts. This design captures key aspects of real-world social interaction, where behavior is shaped jointly by individual identity cues and local social context; detailed theoretical motivations and implementation are provided in subsequent sections.

We evaluate SNSim on multiple real-world social media datasets by comparing simulated outputs with observed data in terms of language use, opinion trending, and network structure. Ablation studies further show that each module plays a distinct role in shaping simulation outcomes.

In summary, our contributions are threefold. First, we provide a clear conceptualization of social network simulation as the joint modeling of language, opinion, and network interactions. Second, we propose SNSim, a modular LLM-based simulation framework that implements this conceptualization through structured prompting. Third, we validate our approach on real-world data, offering empirical evidence for the role of each component. Together, our work advances LLM-based social simulation from heuristic text generation toward a principled methodology for studying complex social dynamics.

2 Framework Overview: SNSim

SNSim is a modular framework for LLM-driven social network simulation. The framework consists of four components: *Initialization*, *Simulation Process*, *Environment*, and an optional *External Control* interface. Initialization constructs the initial agent set and interaction graph from real data. The simulation process iteratively generates content, updates agent states, and evolves network structure. The environment records all intermediate states for analysis and evaluation.

Theoretical grounding. SNSim is grounded in Social Identity Theory (SIT) (Tajfel et al., 1979) and Self-Categorization Theory (SCT) (Maines, 1989) as operational principles for prompt construction and agent evolution. Both theories emphasize that social behavior is jointly shaped by *individual attributes* and *group context*. Guided by this view, SNSim adopts an **individual-community dual-channel** design: each step conditions agents on (i) *personal information* (profile and historical posts) and (ii) *community information* (local group norms and shared posts), consistently across all three modules in simulation process (Figure 1).

2.1 Initialization: Identity Construction and Agent Setup

SNSim initializes from real-world social media data by selecting active users, constructing an interaction graph from observed behaviors (e.g., retweets/replies), and inferring community structure via Louvain clustering (Blondel et al., 2008). Each agent is assigned (i) a community identity and (ii) individual-level attributes derived from historical posts.

Algorithm 1 SNSim Simulation Process

Require: Social graph $G = (V, E)$; initial memory buffers $\{\mathcal{M}_v^0\}_{v \in V}$; profile module $\mathcal{P}$; cognition module $\mathcal{C}$; rewiring probability p; total steps T.
Ensure: Opinion trajectory $\{\mathbf{s}^t\}_{t=1}^T$, generated posts $\mathcal{T}$, evolving graphs $\{G_t\}_{t=1}^T$.

Initialize $G_0 \leftarrow G$; $\mathcal{M}_v \leftarrow \mathcal{M}_v^0$ for all v; $\mathcal{T} \leftarrow \emptyset$
for $t \leftarrow 1$ **to** T **do**
  Sample an active agent $v \sim V$
  # Profile module
  $P_v \leftarrow \textsc{GetProfile}(v, \mathcal{P})$
  $\mathcal{H}_v \leftarrow \textsc{ReadNeighbors}(v, G_{t-1}, \mathcal{T})$
  # Cognition module
  $(\psi_v^{\text{pers}}, \psi_v^{\text{group}}) \leftarrow \textsc{Reflect}(\mathcal{M}_v, \mathcal{H}_v, \mathcal{C})$
  $x_v^t \leftarrow \textsc{LLM\_GenPost}(P_v, \psi_v^{\text{pers}}, \psi_v^{\text{group}})$
  $s_v^t \leftarrow \textsc{Opinion\_Classify}(x_v^t)$
  $\mathcal{T} \leftarrow \mathcal{T} \cup \{x_v^t\}$
  $\mathcal{M}_v \leftarrow \textsc{UpdateMem}(\mathcal{M}_v, \mathcal{H}_v \cup \{x_v^t\})$
  # Rewire module
  if $\textsc{Bernoulli}(p) = 1$ **then**
    $\mathcal{N}_v \leftarrow \textsc{GetNeighbors}(v, G_{t-1})$
    $\mathcal{C}_v \leftarrow \textsc{GetCandidates}(v, G_{t-1})$
    $(u^-, u^+) \leftarrow \textsc{Rewire}(P_v, \psi_v^{\text{pers}}, \psi_v^{\text{group}}, \mathcal{N}_v, \mathcal{C}_v)$
    $G_t \leftarrow \textsc{ApplyEdge}(G_{t-1}, v, u^-, u^+)$
  else
    $G_t \leftarrow G_{t-1}$
  end if
  Record s_v^t into $\mathbf{s}^t$
end for
return $\{\mathbf{s}^t\}_{t=1}^T, \mathcal{T}, \{G_t\}_{t=1}^T$

2.2 Simulation Process

Algorithm 1 summarizes the simulation loop. For clarity, definitions of symbols and prompt templates are provided in Appendices B. At each step, an activated agent reads a bounded local context, constructs a cognitive context via reflection over personal memory and recent neighbor posts, and then generates a post conditioned on **Profile** and **Cognition**. The generated post is classified to update the agent's opinion state, and the agent may optionally adapt its ties via **Rewire**. The algorithm explicitly identifies LLM-dependent operations, enabling transparent analysis of agent behavior and

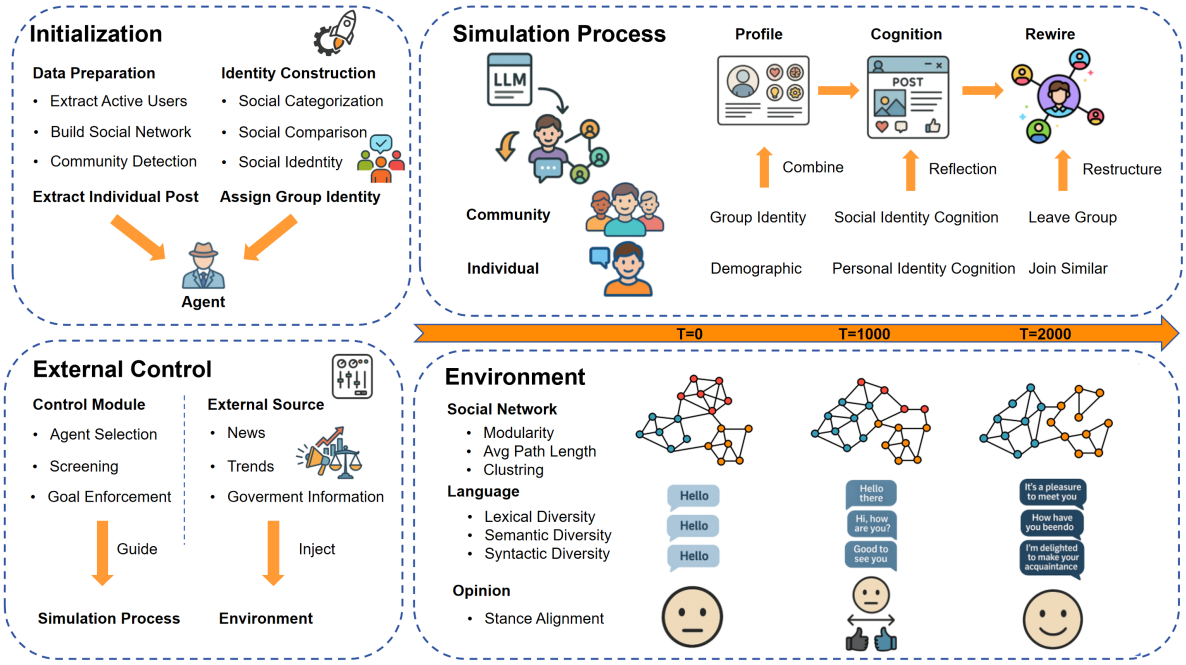

Figure 1: Architecture of the SNSim framework, illustrating the initialization and simulation process, with the environment preserving intermediate states for later evaluation. The framework consists of three core modules—Profile, Cognition, and Rewire—supported by initialization and external control components. The environment continuously logs social dynamics and evaluates network structure, Opinion (Stance Alignment), and language diversity across time.

reproducible simulation.

2.2.1 Profile: Identity Conditioning

The **Profile** module provides identity-conditioned priors for LLM generation by modeling both individual-level and community-level identity information. Individual-level identity captures relatively stable personal attributes, such as demographic characteristics, role identities, and stylistic preferences, inferred from an agent's historical tweets. Community-level identity represents higher-order social identity signals derived from community membership, including prevailing opinion orientations and normative discourse patterns. These identity descriptors are used to calibrate each agent's identity expression in LLM prompts. Details of the profile construction process are provided in Appendix C.

By conditioning prompts on both individual attributes and community identity, SNSim enables agents to generate language that is identity-consistent at the individual level and aligned with group-level norms, reflecting Social Identity Theory (SIT) (Tajfel et al., 1979).

2.2.2 Cognition: Opinion and Posting

The **Cognition** module governs opinion updating and content generation by integrating identity profiles, recently observed neighbor posts, and reflective summaries. Opinion evolution is modeled as a context-dependent self-categorization process within an LLM-driven setting.

Reflection consists of individual reflection over an agent's own posting history and community reflection over observed neighbor discourse. Generated posts are mapped to discrete opinion labels, which serve as the agent's opinion state at each simulation step.

2.2.3 Rewire: Network Adaptation

The **Rewire** module governs network evolution through identity-driven structural decisions informed by agent profiles, recent interactions, and observed community characteristics. Agents preferentially strengthen ties with identity-aligned peers while disengaging from incongruent groups, leading to endogenous community restructuring and echo-chamber formation. Additional details on the framework and prompt templates are provided in Appendix D.

3 Evaluation Metrics

We evaluate simulation fidelity along three aligned axes—**language**, **opinion**, and **network structure**—corresponding to the Profile, Cognition, and Rewire modules.

Language Following prior benchmarking work (Guo et al., 2025), we measure the diversity of generated text at lexical, semantic, and syntactic levels. All metrics are aggregated across agents and time and compared to the distributions observed in real data.

Opinion At each simulation step, agents generate a post which is classified into a discrete stance category using an LLM-based classifier. To evaluate opinion, we compare the stance of each generated post with that of the corresponding real world post by the same agent at the same time step in the dataset. This setup enables direct alignment and comparison between simulated and observed stances over time. Our primary evaluation reports the mean squared error (MSE) between the generated stance trajectories and the ground truth stance curves.

Network structure We assess emergent structural properties of the simulated networks using metrics such as modularity, clustering coefficient, average path length, and assortativity (Newman and Park, 2003; Watts and Strogatz, 1998; Toivonen et al., 2009). These metrics are computed on network snapshots across time and compared against their counterparts observed in the real networks, with discrepancies quantified using MSE. For clarity of presentation, MSE values for network metrics are reported after scaling by a factor of 10^3; this rescaling affects only the numerical scale and does not alter relative comparisons or conclusions.

4 Experiment

4.1 Dataset

We evaluate SNSim on three real world social media datasets covering distinct domains and platforms.

COVID-19 Vaccination Dataset. Collected from English language Twitter using COVID-19-related keywords (Lamsal, 2021), this dataset focuses on vaccine hesitancy and includes tweets labeled as *pro-vaccine*, *anti-vaccine*, or *neutral* (Zaidi et al., 2023).

Ukraine War Dataset. Proposed by (Perera and Karunasekera, 2024), this dataset contains one million English language tweets collected in August 2022. Each tweet is annotated with stance labels (*pro-Russian*, *pro-Ukrainian*, or *neutral*), enabling the study of polarized political discourse.

PolitiSky24: U.S. Election 2024 Bluesky Dataset. Introduced in (Rostami et al., 2025), PolitiSky24 is the first stance detection dataset for the 2024 U.S. presidential election on Bluesky. It provides stance labels derived from an LLM assisted pipeline, along with reasoning explanations and full interaction networks.

For clarity of presentation, most experimental results reported in this paper are averaged across the three datasets. Dataset specific results and additional analyses are provided in the Appendix I.

4.2 Experimental Settings

All simulations were conducted on a server equipped with an **NVIDIA L40s GPU**. We implement SNSim in Python using `llama.cpp` (Gerganov, 2023) for efficient local inference, with agents powered by the open source **Qwen2.5-32B-Instruct** model (Team et al., 2024) (4-bit quantized). This model is chosen for its stable performance in both social content generation and zero-shot stance detection in preliminary experiments. To ensure reproducibility and controllability, we fix the decoding temperature to 0 and run each dataset with three different random seeds. For each simulation run, we execute 2,000 activation steps, corresponding to the generation of 2,000 posts. This scale is sufficient to capture stable opinion trajectories and emergent network dynamics, while avoiding unnecessary computational overhead that would not materially affect the evaluated metrics.

4.3 Baseline Models

We include two categories of reference baselines. First, following common practice in social simulation, we adopt an **equation-based opinion dynamics model** (Sasahara et al., 2021), denoted as *Equation*. Such models are effective at updating numerical opinions but do not generate text, and thus cannot capture language-mediated interactions on social media. Second, we report several recent LLM- or agent-based simulators, including HiSim (Mou et al., 2024), FPS (Liu et al., 2024), and SOD (Chuang et al., 2024), as additional reference points. Since these frameworks typically do not model dynamically evolving network structure and are designed for objectives different from our joint co-evolution setting, we treat them as partially comparable baselines.

To provide a **fair and expressive reference baseline**, we define our main baseline as a *minimal*

LLM-driven simulation. This baseline corresponds to an off-the-shelf LLM agent without any task-specific modular design, reflecting the behavior of a generic LLM-based agent in the absence of structured modeling. Concretely, the baseline receives **exactly the same raw observational inputs** as our full model, including the agent's own historical posts and the posts from neighboring users visible within a fixed context window. However, unlike the full model, these inputs are not explicitly organized into structured components such as identity profiles or reflective cognition processes; instead, all agent behavior—including content generation and opinion prediction—is driven directly by the LLM under a uniform prompting scheme.The prompt template used by the base model is provided in Appendix E.

5 Experimental Results

Table 1 summarizes the overall performance of SNSim across three complementary dimensions of social simulation: language generation, opinion dynamics, and network structure. The table is organized by dimension, with each block reporting metrics aligned with one aspect of the simulation process, while columns correspond to different classes of baseline methods with distinct modeling assumptions. Our primary comparison is between the **Base Model** and the **SNSim**. Both models receive *exactly the same observational inputs* and differ only in whether this information is explicitly organized and operationalized through the Profile, Cognition, and Rewire modules. Equation-based models and existing LLM-driven simulators are included for reference and are evaluated only on the dimensions supported by their respective modeling capabilities.

Overall, SNSim consistently improves simulation fidelity across all three dimensions. Across language, opinion, and network metrics, SNSim is the only approach that demonstrates systematic improvements across heterogeneous evaluation criteria, highlighting the benefit of jointly modeling linguistic behavior, opinion dynamics, and network evolution within a unified framework.

Language. Language metrics are reported only for models that generate textual content; equation-based methods are therefore excluded from this evaluation. Among LLM-driven baselines (FPS, SOD, and HiSim), generation may condition on some form of user context or profile information. However, such representations are typically treated in a coarse or undifferentiated manner and are not explicitly decomposed into individual-level attributes and community-level identity signals. As a result, these models have limited capacity to regulate identity-consistent language use across social contexts and do not account for the interaction between language generation and network evolution (i.e., rewiring).

In contrast, the Base Model and SNSim form a strictly controlled comparison pair. They receive exactly the same inputs and share overlapping prompt components, differing only in whether structured Profile, Cognition, and Rewire modules are used to organize and utilize this information. Relative to the Base Model, SNSim improves lexical diversity by 37%, with additional gains of 14% and 10% in semantic and syntactic diversity, respectively. SNSim achieves the best performance among LLM-driven methods on lexical and syntactic diversity, and competitive performance on semantic diversity, indicating that explicitly modeling identity and context leads to more varied and socially grounded language generation.

Opinion. We operationalize opinion as discrete stance states (FAVOR / AGAINST / NONE) inferred from each generated post. Opinion dynamics are evaluated using aligned stance trajectories for all methods, including equation-based and LLM-driven approaches. Lower MSE and MAE indicate closer numerical alignment with ground-truth temporal evolution, while the coefficient of determination (R^2) reflects the extent to which a model explains variance beyond a constant baseline.

Compared with the Base Model, SNSim reduces stance MSE and MAE by 15% and 9%, respectively. The R^2 score also improves by 12%, indicating a better fit to the temporal structure of observed opinion dynamics, although the overall R^2 remains negative. This result suggests that while stance prediction remains challenging, explicitly organizing identity and contextual information contributes to more coherent opinion trajectories over time.

Network Structure. Network metrics are reported only for models that explicitly simulate network evolution. These metrics quantify deviation from observed network statistics rather than edge-level prediction error, with lower values indicating better structural alignment.

SNSim substantially reduces error on several key structural metrics, including a 73% reduction

Category	Metric	Equation	FPS	SOD	HiSim	Base Model	SNSim	Improve.
Language	Lexical ↑	–	0.025	0.012	0.036	0.043	**0.059***	37%
	Semantic ↑	–	0.347	0.233	**0.597**	0.478	0.544*	14%
	Syntactic ↑	–	0.341	0.349	0.300	0.367	**0.402***	10%
Opinion	Stance MSE ↓	0.722	0.698	0.576	0.554	0.639	**0.541***	15%
	Stance MAE ↓	0.613	0.615	0.535	0.499	0.539	**0.493***	9%
	Stance R^2 ↑	-1.910	−1.947	−1.541	−1.419	−1.532	**-1.354***	12%
Network	Modularity ↓	0.454	–	–	–	0.641	**0.173***	73%
	Path Length ↓	119.365	–	–	–	176.768	**72.449***	59%
	Clustering ↓	0.069	–	–	–	0.079	**0.064***	19%
	Assortativity ↓	0.567	–	–	–	**0.308**	0.340*	-10%

Table 1: Overall comparison between the Base Model and SNSim framework. All metrics report average performance across three datasets. Arrows (↑ / ↓) indicate whether higher or lower values correspond to better performance. * indicates a statistically significant difference between SNSim and the Base Model ($p < 0.05$).

in modularity error and a 59% reduction in average path length error relative to the Base Model. The clustering coefficient error is further reduced by 19%, indicating improved recovery of local connectivity patterns. In contrast, assortativity error increases slightly (10%), reflecting the known sensitivity of assortativity to fine-grained degree correlations and local mixing patterns. This suggests that while SNSim effectively captures meso- and macro-scale network organization, modeling micro-level homophily remains a more challenging aspect of network evolution.

Taken together, these results show that SNSim achieves the strongest or near-strongest performance across language, opinion, and network dimensions under a unified experimental setting. The **Improvement** column in Table 1 quantifies the contribution of structured agent modeling under otherwise identical input conditions. These findings underscore the importance of jointly modeling language generation, opinion dynamics, and network evolution, as improvements along any single dimension alone are insufficient to reproduce the coupled processes observed in real-world social systems.

5.1 Ablation Study

Table 2 and Figure 2 present an ablation study that disentangles the contributions of the Profile (P), Cognition (C), and Rewire (R) modules in SNSim. While Table 2 reports aggregated performance across three high-level dimensions, Figure 2 further visualizes the relative improvements of individual metrics with respect to the baseline.

From a network perspective, enabling the Rewire module plays a dominant role in improving structural fidelity. As shown in Figure 2, Rewire yields

Config.	Network↓	Stance↓	LangDiv↑
Baseline	0.044	0.639	0.296
+Profile	0.044	0.539	**0.362**
+Cognition	0.044	0.586	0.358
+C+P	0.044	**0.527**	**0.362**
Full	**0.018**	0.541	0.335

Table 2: Ablation study of SNSim components. Lower Network and Stance indicate better performance; higher LangDiv indicates greater linguistic diversity.

substantial gains in modularity and path length (over 70% and 50% relative improvement, respectively), whereas Profile and Cognition alone have negligible or inconsistent effects on these metrics. This pattern is reflected in Table 2, where only the full configuration significantly reduces the aggregated network error, confirming that structural adaptation is essential for reproducing realistic network evolution.

In contrast, stance prediction benefits primarily from Profile and Cognition. Both modules independently reduce stance error, and their combination (C+P) achieves the best overall stance performance among all configurations. Figure 2 shows consistent positive contributions from Profile and Cognition across stance-related metrics, while Rewire introduces a mild degradation when considered in isolation. This suggests that identity conditioning and memory-based cognition provide strong signals for stance consistency, whereas dynamic rewiring may introduce additional variability in opinion trajectories.

Language diversity exhibits a complementary pattern. Profile and Cognition consistently improve lexical, semantic, and syntactic diversity, as shown by positive gains across all language related metrics in Figure 2. Their combination yields the high-

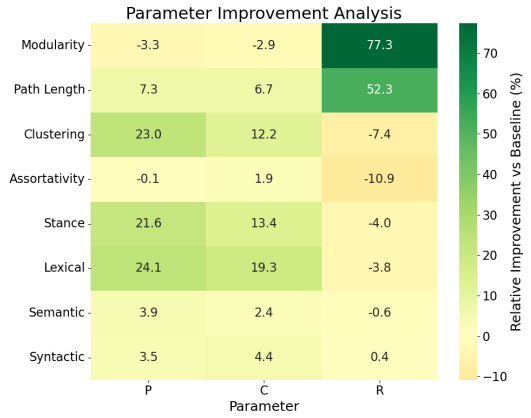

Figure 2: Ablation study of individual parameters (P, C, R). Values denote relative improvement over the Base Model across evaluation metrics.

est aggregated language diversity score in Table 2. Although Rewire alone does not directly enhance linguistic diversity, the full model maintains a comparably high level of language diversity while simultaneously achieving substantial improvements in network structure.

Overall, the ablation study reveals a clear division of labor among the components: Profile and Cognition primarily enhance stance coherence and linguistic diversity, whereas Rewire is indispensable for capturing realistic network dynamics. The full SNSim configuration integrates these complementary effects, achieving the best balance across structural, attitudinal, and linguistic dimensions. Due to space constraints, we report only the main ablation results here. Additional ablation studies that further analyze each module in isolation are provided in Appendix G.

6 Case Study: Covid-19 Vaccination

6.1 Language Diversity

Table 3 provides a qualitative comparison between the baseline configuration and the full model on two representative users (Case #35 and #42), both of whom exhibit *skeptical, anti-vaccine tendencies in the ground-truth data*.

The baseline model consistently generates safe, generic, and assistant-like content aligned with mainstream pro-vaccine narratives. Across both cases, its outputs emphasize neutral scientific consensus (e.g., "reliable evidence," "public health") with little variation in tone or stance. Consequently, the generated texts for Case #35 and #42 are highly similar, indicating a failure to reflect the users' ground-truth anti-vaccine orientations or to differentiate individual personas.

In contrast, the full model produces persona-consistent outputs that better align with the users' skeptical ground-truth stances. For Case #35, it adopts rights-based framing and group-specific rhetoric (e.g., "Big Pharma," "choice"), while for Case #42 it employs emotionally charged language and stylistic markers characteristic of vaccine-hesitant discourse. These distinctions reflect improved alignment with user-specific opinions and expressive styles.

Overall, the comparison shows that the full model preserves opinion diversity and stylistic variation grounded in observed user behavior, whereas the baseline collapses toward homogeneous, safety-oriented responses. This capability is critical for social network simulation, where faithfully modeling non-mainstream and minority narratives is essential for downstream analysis.

6.2 Opinion Trending Analysis

We analyze opinion dynamics on the COVID-19 dataset by comparing the evolution of average stance scores over simulation steps. Figure 3 contrasts the ground-truth trajectory with outputs from the baseline model and SNSim.

The baseline model produces an overly stable, consistently pro-vaccine trajectory. Although aligned with mainstream public health narratives, it fails to capture the temporal fluctuations and polarization evident in the ground truth, remaining largely insensitive to periods of increased skepticism. In contrast, SNSim more closely follows the ground-truth trend, capturing both the overall direction and the magnitude of stance changes over time. In particular, when the ground truth exhibits sharp drops or recoveries in average opinion, SNSim responds with corresponding variations, whereas the baseline remains overly smooth and biased toward a single stance.

These results demonstrate that SNSim better preserves temporal stance dynamics in social media discourse, avoiding collapse to static or safety-oriented opinions and instead reflecting the evolving and sometimes volatile nature of real-world COVID-19 discussions.

6.3 Network Evolution

To evaluate micro-level structural fidelity, we visualize the evolution of a 3-hop ego-network for a highly active user (Figure 4), aligning the

Table 3: Case Study: Comparison of Generated Tweets between Base Model and SNSim Model

Case ID	Base Model	SNSim
#35	It's important to base our decisions on reliable scientific evidence. Vaccines have saved countless lives and continue to be a crucial tool in public health. Let's support factual information and responsible health policies. *(Stance: FAVOR)*	The government and big pharma are pushing a narrative that doesn't align with the truth. We must stand firm for our right to choose what enters our bodies. #VaccineChoice #PersonalFreedom *(Stance: AGAINST)*
#42	It's important to base our decisions on reliable scientific evidence. While concerns about new territories in medical research are valid, vaccines have saved countless lives and eradicated diseases. Let's ensure we're informed from credible sources. *(Stance: FAVOR)*	The more they try to control, the more we must question. #vaccinerisks aren't just about needles; they're about our right to choose and live freely. #individualfreedom #questionauthority *(Stance: AGAINST)*

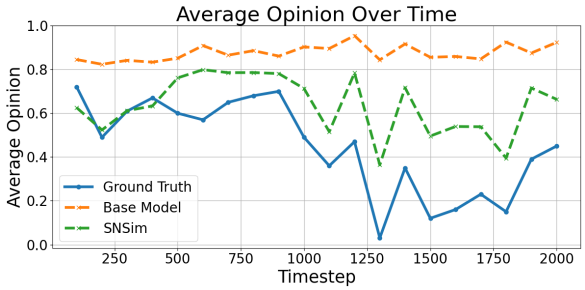

Figure 3: Average opinion over time on the COVID-19 dataset. We compare the ground-truth opinion trajectory with the baseline model and SNSim.

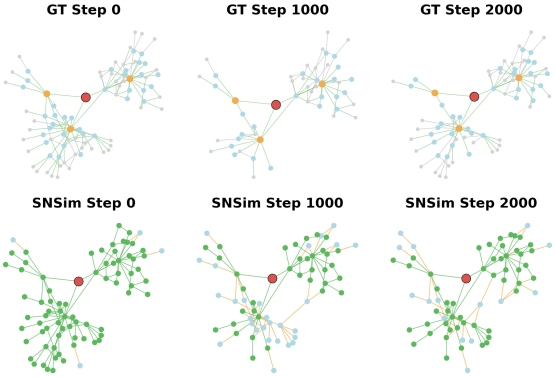

Figure 4: Spatiotemporal evolution of a representative user's 3-hop ego-network in the COVID-19 dataset. The top row shows the Ground Truth (GT) network at $t = 0, 1000, 2000$, and the bottom row shows the corresponding SNSim simulations. Nodes are colored by distance from the ego, and green nodes/edges indicate correctly predicted connections.

Ground Truth (GT) and simulated networks at three timesteps ($t = 0, 1000, 2000$).

At $t = 0$, the ego resides in a locally dense community with a clear separation between immediate neighbors and peripheral nodes. SNSim closely reconstructs this initial topology, preserving both direct and higher-order neighborhood structure. As the network evolves, the simulated snapshots remain well aligned with the GT, with most local connections correctly predicted, as indicated by the prevalence of green edges.

Even when deviations occur, SNSim-generated links tend to follow plausible local patterns, such as intra-community attachment, rather than forming arbitrary long-range connections. This suggests that SNSim captures not only static neighborhood structure but also realistic mechanisms of local network evolution.

7 Conclusion

We introduced **SNSim**, a theory-grounded and modular framework for LLM-based social network simulation that jointly models **language**, **opinion**, and **network structure**. By operationalizing Social Identity Theory and Self-Categorization Theory through an individual-community dual-channel prompt design, SNSim enables agents' expression, opinion evolution, and structural adaptation to co-evolve in an interpretable simulation loop. Experiments on three real-world datasets show that SNSim consistently outperforms a minimal LLM baseline across linguistic diversity, stance trajectory alignment, and network structural fidelity. Ablation and case studies further reveal a clear division of labor: Profile and Cognition primarily enhance stance coherence and linguistic realism, while Rewire is essential for capturing realistic network dynamics. Together, these results demonstrate that theory-aligned prompt structuring provides a principled path beyond heuristic text generation for LLM-based social simulation.

8 Limitations

SNSim relies on structured prompting and LLM-driven decisions, and therefore inevitably inherits known sensitivities of large language models, in-

cluding dependence on prompt phrasing, domain shifts, and normative or safety-aligned priors that may introduce systematic ideological biases (e.g., overly mainstream or polarized stances). While identity conditioning and reflective cognition partially mitigate these effects, prompt design alone cannot fully eliminate model-internal biases. In addition, opinions are evaluated using discrete stance labels, which are compatible with existing datasets but may oversimplify nuanced opinion states (e.g., ambivalence or conditional support) and allow classification errors to propagate into memory and rewiring decisions; richer representations and evaluation metrics remain to be explored. Furthermore, the Rewire module abstracts network adaptation through bounded, identity-aware tie changes and does not explicitly model platform-specific mechanisms such as recommendation systems, visibility constraints, or moderation, nor regimes with strongly asymmetric or density-varying network dynamics. Finally, as with any LLM-based social simulation, generated content may reproduce or amplify biased or harmful narratives if misused, underscoring the need for careful governance and responsible deployment.

References

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Yun-Shiuan Chuang, Agam Goyal, Nikunj Harlalka, Siddharth Suresh, Robert Hawkins, Sijia Yang, Dhavan Shah, Junjie Hu, and Timothy Rogers. 2024. Simulating opinion dynamics with networks of llm-based agents. In *Findings of the association for computational linguistics: NAACL 2024*, pages 3326–3346.

Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98.

Noah E Friedkin and Eugene C Johnsen. 1990. Social influence and opinions. *Journal of mathematical sociology*, 15(3-4):193–206.

Georgi Gerganov. 2023. llama.cpp: Port of facebook's llama model in c/c++. https://github.com/ggerganov/llama.cpp.

Ilker Gül, Rémi Lebret, and Karl Aberer. 2024. Stance detection on social media with fine-tuned large language models. *arXiv preprint arXiv:2404.12171*.

Yanzhu Guo, Guokan Shang, and Chloé Clavel. 2025. Benchmarking linguistic diversity of large language models. *Transactions of the Association for Computational Linguistics*, 13:1507–1526.

Rabindra Lamsal. 2021. Design and analysis of a large-scale covid-19 tweets dataset. *Applied Intelligence*, 51(5):2790–2804.

Yuhan Liu, Xiuying Chen, Xiaoqing Zhang, Xing Gao, Ji Zhang, and Rui Yan. 2024. From skepticism to acceptance: Simulating the attitude dynamics toward fake news. *arXiv preprint arXiv:2403.09498*.

David R Maines. 1989. Rediscovering the social group: A self-categorization theory.

Xinyi Mou, Zhongyu Wei, and Xuan-Jing Huang. 2024. Unveiling the truth and facilitating change: Towards agent-based large-scale social movement simulation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 4789–4809.

Mark EJ Newman and Juyong Park. 2003. Why social networks are different from other types of networks. *Physical review E*, 68(3):036122.

Marios Papachristou and Yuan Yuan. 2025. Network formation and dynamics among multi-llms. *PNAS nexus*, 4(12):317.

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*, pages 1–18.

Kushani Perera and Shanika Karunasekera. 2024. Quantifying opinion rejection: A method to detect social media echo chambers. In *PAKDD*, pages 57–69. Springer.

Peyman Rostami, Vahid Rahimzadeh, Ali Adibi, and Azadeh Shakery. 2025. Politisky24: Us political bluesky dataset with user stance labels. *arXiv preprint arXiv:2506.07606*.

Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2021. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, 4(1):381–402.

Yanhui Sun, Wu Liu, Wentao Wang, Hantao Yao, Jiebo Luo, and Yongdong Zhang. 2025. Dynamix: Large-scale dynamic social network simulator. *arXiv preprint arXiv:2507.19929*.

Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An integrative theory of intergroup conflict. *Organizational identity: A reader*, 56(65):9780203505984–16.

Qwen Team and 1 others. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2(3).

Riitta Toivonen, Lauri Kovanen, Mikko Kivelä, Jukka-Pekka Onnela, Jari Saramäki, and Kimmo Kaski. 2009. A comparative study of social network models: Network evolution models and nodal attribute models. *Social networks*, 31(4):240–254.

Chenxi Wang, Zongfang Liu, Dequan Yang, and Xiuying Chen. 2025. Decoding echo chambers: Llm-powered simulations revealing polarization in social networks. In *Proceedings of the 31st international conference on computational linguistics*, pages 3913–3923.

Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

Zainab Zaidi, Mengbin Ye, Fergus Samon, Abdisalan Jama, Binduja Gopalakrishnan, Chenhao Gu, Shanika Karunasekera, Jamie Evans, and Yoshihisa Kashima. 2023. Topics in antivax and provax discourse: yearlong synoptic study of covid-19 vaccine tweets. *Journal of Medical Internet Research*, 25:e45069.

Wenzhen Zheng and Xijin Tang. 2024. Simulating social network with llm agents: An analysis of information propagation and echo chambers. In *International Symposium on Knowledge and Systems Sciences*, pages 63–77. Springer.

Xuhui Zhou, Zhe Su, Tiwalayo Eisape, Hyunwoo Kim, and Maarten Sap. 2024. Is this the real life? is this just fantasy? the misleading success of simulating social interactions with llms. *arXiv preprint arXiv:2403.05020*.

A Related Work

A.1 Social Simulation with Language Models

Recent advances in LLMs, such as GPT-4 and other autoregressive models, have enabled novel applications in simulating human behavior and discourse in online platforms. Prior work has used LLMs to generate synthetic tweets (Park et al., 2022), simulate opinion dynamics (Wang et al., 2025), and study political polarization (Zheng and Tang, 2024). However, these models often treat users as isolated text emitters, lacking personalization, memory, or social context. More recently,DYNAMIX (Sun et al., 2025), a large-scale dynamic social network simulator that explicitly models the co-evolution of opinions and network structure. However, the absence of released prompts and code makes reproducible evaluation and direct comparison with LLM-driven social simulation frameworks challenging.

A.2 Network Dynamics and Opinion Clustering

Classical opinion dynamics models, such as the Bounded Confidence Model (Deffuant et al., 2000) and Friedkin-Johnson Dynamics Model (Friedkin and Johnsen, 1990), simulate influence spread and belief update on fixed or evolving networks. Rewiring mechanisms have also been studied to reflect echo chambers and homophily (Sasahara et al., 2021). Our Rewire module draws inspiration from this literature but integrates it with LLM-generated content to drive connection updates in a more data-driven and content-aware way.

A.3 Multi-faceted Evaluation of Simulation Fidelity

Evaluating simulation realism requires considering multiple axes: language diversity (Guo et al., 2025), stance alignment (Gül et al., 2024), and network structure (Papachristou and Yuan, 2025). Existing benchmarks typically evaluate only one aspect in isolation. We adopt a comprehensive evaluation protocol, measuring SNSim's performance across all three axes using quantitative metrics and human validation where applicable.

B Notation Used in Algorithm 1

This section defines the symbols and variables used in Algorithm 1 for clarity and reference.

Agents and graph. We denote the social graph by $G = (V, E)$, where V is the set of agents (users) and E represents directed social ties (e.g., follow or interaction links). At simulation step t, the evolving graph state is denoted by G_t. An active agent selected at step t is denoted by $v \in V$.

Memory and observed context. Each agent v maintains a bounded memory buffer $\mathcal{M}_v$, which stores recent interaction records, including generated posts and observed neighbor content. The initial memory buffer is denoted by $\mathcal{M}_v^0$. At step t, $\mathcal{H}_v$ denotes the set of posts currently observed by agent v from its neighbors in G_{t-1}, corresponding to the local screen context read via READNEIGHBORS.

Cognitive states. The Cognition module produces two intermediate cognitive representations for the active agent v: a personal cognitive state ψ_v^{pers}, reflecting individual-level experience and preferences derived from $\mathcal{M}_v$, and a group-aware cognitive state ψ_v^{group}, capturing salient signals from the agent's social neighborhood $\mathcal{H}_v$. These representations are generated by the REFLECT operation and are expressed as natural-language summaries.

Generated content and opinion states. The post generated by agent v at step t is denoted by x_v^t, obtained via the LLM_GENPOST operation conditioned on the agent profile and cognitive states. Each generated post is subsequently classified into a discrete stance label $s_v^t \in \{\text{FAVOR}, \text{AGAINST}, \text{NONE}\}$ using the OPINION_CLASSIFY operation. We denote by $\mathbf{s}^t$ the vector of opinion states across all agents at step t, which records the instantaneous opinion configuration of the simulated system.

Generated post collection. The set $\mathcal{T}$ denotes the collection of all posts generated during the simulation up to the current step. This set grows monotonically as new posts x_v^t are produced and is used both for updating agent memory and for constructing neighbor observation contexts in subsequent steps.

Rewiring-related notation. For a given agent v, $\mathcal{N}_v$ denotes the set of its current neighbors in graph G_{t-1}, while $\mathcal{C}_v$ denotes a bounded set of candidate agents considered for potential new connections. The rewiring operation returns a pair (u^-, u^+), where u^- is the neighbor selected for unfollowing and u^+ is the candidate selected for following. Applying this operation yields the updated graph state G_t.

Simulation parameters. The total number of simulation steps is denoted by T. At each step, rewiring is activated with probability p via a Bernoulli trial. The full simulation outputs include the opinion trajectory $\{\mathbf{s}^t\}_{t=1}^T$, the generated post set $\mathcal{T}$, and the sequence of evolving graphs $\{G_t\}_{t=1}^T$.

The prompt templates corresponding to the LLM-based operations used in Algorithm 1 are documented in Appendix F.

C Profile Construction via LLM-Based Inference

This appendix provides detailed documentation of the profile construction process in SNSim, including prompt templates, inference stages, and illustrative examples. The goal of this appendix is to improve transparency and reproducibility by explicitly describing how individual- and community-level identity signals are inferred from textual and network data using large language models.

C.1 Overview

SNSim constructs agent profiles through a multi-stage LLM-based inference pipeline that integrates individual posting behavior with community-level social identity signals. Rather than assuming predefined demographic or ideological attributes, SNSim infers profiles directly from observed language use and network structure. This design allows agents to exhibit heterogeneous yet socially grounded identities that evolve naturally from data.

The profile construction pipeline consists of four stages: (1) individual-level demographic inference, (2) structural community detection, (3) community-level identity discovery and abstraction, and (4) profile refinement within a unified identity space.

C.2 Individual-Level Profile Inference

For each user, SNSim aggregates historical posts by merging both original authorship (`UserId`) and retweet-source identities (`SourceUserId`), treating them as a unified expressive footprint. This ensures that both content production and amplification behavior contribute to profile inference.

A large language model is then used to infer coarse-grained demographic and background attributes from each user's historical posts. The in-

ferred attributes include age group, gender, education level, occupation field, and cultural or ethnic identity. These attributes are treated as *soft priors* for conditioning downstream generation rather than as factual annotations.

To avoid overconfident or speculative inference, the model is explicitly instructed to output "Uncertain" when insufficient evidence is present.

C.2.1 Individual-Level Inference Prompt

Individual-Level Profile Inference Prompt

Below is a sample of tweets from a user identified as `{user_id}`.
Based on the user's tweet content, language patterns, and topics, please infer their likely demographic attributes. Do not summarize or quote the tweets directly. If an attribute is difficult to infer, write "Uncertain".
Name: `{user_id}`
Attributes to infer:

1. Age group (e.g., 18–24, 25–34, 35–44, 45–60, 60+)
2. Gender (e.g., Male, Female, Nonbinary, Uncertain)
3. Education level (e.g., High school, Bachelor's, Postgraduate, Uncertain)
4. Occupation field (e.g., Healthcare, Education, IT, Student, Unemployed, Uncertain)
5. Cultural or ethnic identity (e.g., South Asian background, African-American identity, Uncertain)

Selected tweets:

- {tweet_1}
- {tweet_2}
- ...

C.2.2 Illustrative User-Level Example

As an illustrative example, for an anonymized user (*User U–XXX*), the model infers an age range of 25–34 and a postgraduate education level, based on sustained engagement with public health policy, vaccine efficacy, and international health initiatives. Gender and cultural identity are marked as "Uncertain" due to the absence of explicit linguistic cues.

Beyond demographic attributes, the same user's content is characterized as analytical in communication style, scientifically grounded in belief orientation, and strongly aligned with collective public health values. This example illustrates how SNSim captures interpretable identity tendencies without requiring explicit self-disclosure from users.

C.3 Community Detection and Social Group Formation

To capture higher-order social identity, SNSim performs community detection over the retweet network using the Louvain algorithm. The resulting communities represent structurally cohesive social groups and serve as proxies for shared discourse environments, exposure patterns, and latent social identity.

Each user is assigned a community label, which is subsequently used for community-level identity analysis and aggregation.

C.4 Community-Level Identity Discovery

To identify distinguishing identity characteristics across communities, SNSim adopts a contrastive prompting strategy inspired by Social Identity Theory (SIT). For each comparison, two communities are randomly sampled, along with multiple users and tweets from each group. The language model is prompted to explicitly contrast the groups in terms of communication patterns, values, beliefs, and perspectives.

This process is repeated across multiple random samples to reduce sensitivity to individual users or specific tweets.

C.4.1 Community Comparison Prompt

Community-Level Identity Comparison Prompt

The following are tweets written by users from two different social groups. These users have been clustered based on structural communities in a social network.
Please analyze the tweets and identify distinguishing identity-related attributes for each group. Focus on features that reflect their social identity in line with Social Identity Theory (SIT), including communication patterns, values, beliefs, and perspectives.
After describing the characteristics of each group, please summarize the **five most informative attributes** that would be most effective for user profiling in a social simulation model.

C.4.2 Illustrative Community Comparison

In a typical comparison, one community may be characterized by direct and assertive communication, frequent references to collective values such as inclusivity and social justice, and a progressive orientation toward societal change. In contrast, another community may exhibit more individual-focused narratives, lower emotional expressiveness, and a pragmatic or neutral stance toward group-level issues.

Such contrasts consistently emerge across repeated comparisons and form the basis for abstracting higher-level identity dimensions.

C.5 Identity Dimension Abstraction

Aggregating identity-related attributes across multiple community comparisons, SNSim abstracts a compact set of identity dimensions. As an illustrative example, in the Bluesky dataset used in this study, the model derives five conceptually distinct dimensions: *Communication Style*, *Value Orientation*, *Belief Systems*, *Engagement Patterns*, and *Perspective on Community*. The specific dimensions are data-dependent and may vary across datasets, while the abstraction procedure remains unchanged.

C.6 Final Profile Refinement and Usage

In the final stage, each user's profile is refined by re-inferring attributes along the abstracted identity dimensions using the user's historical posts. This ensures that all agents are represented within a unified identity space.

Community-level profiles are constructed by aggregating refined user profiles within the same structural community. These profiles capture shared norms, dominant value orientations, and characteristic engagement patterns, and are used to condition content generation, opinion updating, and network rewiring throughout the SNSim simulation loop.

D Framework Detail

In the previous section, we described how SNSim constructs agent profiles via LLM-based inference, capturing relatively stable individual attributes and socially grounded identity cues (Appendix C). Building on these profiles, this section details how agents *reason*, *act*, and *adapt their social connections* during the simulation.

Specifically, we introduce the design of the **Cognition** and **Rewire** modules. The Cognition module governs how agents interpret observed content and generate new messages, while the Rewire module determines how agents update their social ties in response to evolving discourse. Together, these modules enable SNSim to model not only what agents say, but also how their opinions and social neighborhoods co-evolve over time.

D.1 Cognition Module: Memory- and Profile-Conditioned Reasoning

The Cognition module defines how an agent *perceives*, *interprets*, and *responds* to the surrounding discourse at each simulation step. Rather than relying on numerical update rules, SNSim implements Cognition as an LLM-driven reasoning process that transforms observable social signals into both textual actions and stance updates.

At each activation step, the Cognition module conditions the agent's behavior on three types of information: (i) the agent's current screen exposure (i.e., the set of visible posts), (ii) optional profile-based role and identity priors inferred offline, and (iii) optional contextual signals derived from recent interaction history. These inputs are jointly used to generate the agent's next message and to update its internal opinion state.

Cognition as perception → reasoning → action. Cognition in SNSim follows an explicit three-stage pipeline that maps observable social signals to agent actions. This design mirrors the simulation

loop and makes the agent decision process transparent and auditable.

(1) Perception. At time step t, an activated agent u first *perceives* its local information environment. Concretely, the social media environment returns a bounded "screen" consisting of recent posts from accounts followed by u, together with the agent's own recent posting history. The screen represents the agent's immediate exposure at t and serves as the primary evidence for downstream reasoning.

(2) Reasoning. Given the perceived screen, SNSim constructs a compact *reasoning context* by composing multiple information channels. First, the textual content of the screen provides exposure evidence reflecting the agent's current information environment. Second, when `use_profile=True`, the agent's profile supplies relatively stable role and identity priors, constraining how content is interpreted and expressed. Third, when `use_memory=True`, short-term contextual signals derived from recent interactions are incorporated to enrich the reasoning context.

Memory construction as contextual signals. When memory is enabled, these contextual signals are constructed from two sources: (i) **personal memory**, summarizing the agent's own recent posts stored in `memory_df` into a concise, identity-aware narrative using `SUMMARIZATION_PROMPT_SIT`; and (ii) **group experience**, summarizing the most recent posts of the agent's current neighbors (friends) using `GROUP_HISTORY_SUMMARY_PROMPT_SIT`. Optionally, these summaries can be fused into a brief reflective statement that captures higher-level attitudes shaped jointly by individual experience and perceived group identity. Importantly, memory is treated as a lightweight, natural-language context rather than a persistent hidden state, allowing it to inform reasoning without dominating agent behavior.

Memory modes and controllable ablations. For analysis purposes, SNSim exposes a `memory_mode` switch that controls how (and whether) memory signals are injected into the reasoning process:

- `personal`: include only the personal-history summary.
- `group`: include only the group-experience summary derived from friends' latest posts.
- `raw_friends`: inject friends' latest posts directly without summarization.
- `combined`: include personal summary, group summary, and a short reflective memory (default).

This design enables controlled ablations to disentangle the effects of individual recall, community-conditioned identity, raw exposure to neighbor content, and reflective integration.

(3) Action. Finally, the agent produces an *action* as natural language. The main Cognition routine (`update_opinion_llm_compose`) constructs a structured prompt that includes: (1) an agent identifier, (2) an optional role or profile description, (3) optional memory-derived context determined by `use_memory` and `memory_mode`, and (4) the current screen content. The language model is instructed to generate exactly one message in a constrained format (`New Tweet: <tweet>`), ensuring consistent parsing and downstream processing.

Stance inference and opinion update. After a message is generated, SNSim applies a separate stance inference prompt to classify the text into one of three labels: FAVOR, AGAINST, or NONE, with respect to the task-specific target (e.g., COVID-19 vaccination). The agent's internal opinion state is then updated by mapping these labels to a discrete scalar representation: FAVOR $\rightarrow 1$, AGAINST $\rightarrow -1$, NONE $\rightarrow 0$. This decoupling of language generation from opinion state tracking yields both expressive textual actions and interpretable attitude dynamics, which are subsequently used for evaluation and, when enabled, for network rewiring.

Reproducibility and logging. To support reproducibility and analysis, the simulator logs all LLM outputs at each step (`llm_results.csv`) together with the generated message stream. Moreover, the activation order of agents is externally controlled by iterating through an observed dataset, ensuring one-to-one temporal alignment between simulated actions and ground-truth trajectories during evaluation.

D.2 Rewire Module: LLM-Guided Connection Updates

The Rewire module models the structural adaptation of the social graph, i.e., how agents update their follow relationships as discourse evolves.

While classical models typically rely on probabilistic rewiring mechanisms, SNSim implements rewiring as a content-aware decision process, optionally guided by large language models. This design allows connection updates to be grounded in observed discourse and social identity cues rather than purely stochastic rules.

When rewiring is triggered. In the underlying echo chamber model, rewiring is conceptually governed by a rewiring probability that determines whether an agent updates its connections at a given step. However, in our experimental setting, SNSim leverages observed interaction logs to simulate rewiring in a more deterministic and data-aligned manner.

At each time step, the simulator reads the corresponding row from the observed interaction log. If the row corresponds to a repost or retweet event (`retweet_id` is not null), the simulator treats the step as a reposting action and disables rewiring at that step. If the row corresponds to an original post (`retweet_id` is null) and `use_rewiring=True`, the simulator invokes the agent's rewiring routine and applies the resulting update to the graph. This design enables SNSim to faithfully mirror when structural changes are likely to occur in real data, while retaining compatibility with probabilistic rewiring formulations.

Candidate construction. Given an active agent u, the Rewire module constructs two sets of candidates:

- **Current friends**: the out-neighbors of u in the directed graph, corresponding to accounts that u currently follows. For each friend, the module retrieves the most recent post from their history and formats it as `Friend <id>: <text>`.

- **Follow candidates**: a small set of users not currently followed by u. The current implementation selects a fixed-size subset (e.g., 10 users) from the non-friend user list and retrieves each candidate's most recent post, formatted as `Candidate <id>: <text>`.

If either friends' latest posts or candidate posts are unavailable (e.g., due to empty histories), the rewiring action is skipped for that step to avoid ill-posed decisions.

LLM-based rewiring decision. When LLM-guided rewiring is enabled, the main routine `decide_to_rewire_llm_compose` constructs an instruction prompt that integrates: (1) an optional agent role or identity description derived from the Profile module, (2) optional personal and reflective memory summaries provided by the Cognition module, (3) the latest posts from current friends and candidate users, and (4) an explicit output constraint requiring a `Reason`, `Unfollow_ID`, and `Follow_ID`. The prompt emphasizes that connection decisions should be informed by content alignment, perceived group membership, and shared values, rather than by numerical opinion distance alone. This allows rewiring to reflect socially grounded judgments that are difficult to capture with hand-crafted rules.

Applying and logging structural updates. If both `unfollow_id` and `follow_id` are successfully extracted from the LLM output, the simulator applies the update by calling `social_media.rewire_users(user_id, unfollow_id, follow_id)` on the directed graph. All rewiring decisions are recorded in `rewire_results.csv`, including the full prompt and raw LLM response. These logs enable auditing, reproducibility, and fine-grained error analysis (e.g., invalid identifiers, format violations, or inconsistent rationales).

Non-LLM rewiring baselines. As a point of comparison, SNSim also implements a non-LLM rewiring strategy following prior work on echo chamber dynamics (Sasahara et al., 2021). This heuristic routine (`decide_to_rewire`) unfollows one discordant friend and selects a new account to follow using one of three strategies: `Random` (uniform random selection), `Repost` (friends-of-friends derived from concordant repost sources), or `Recommendation` (rule-based similarity recommendations). When `use_rewiring=False` or when LLM-based decisions are disabled, this heuristic serves as the baseline mechanism for structural evolution.

Summary. Together with the Cognition module, Rewire enables SNSim to model the co-evolution of discourse and network structure. While Cognition governs how agents transform exposure into language and stance updates, Rewire governs how these evolving attitudes reshape the social graph. Both modules produce explicit, auditable interme-

diate artifacts, supporting controlled ablations and reproducible evaluation of language generation, opinion dynamics, and emergent network structure.

D.3 Environment

The environment records all simulation states, including generated posts, opinion trajectories, and evolving network snapshots. These records support downstream analysis of individual behavior and emergent collective dynamics.

D.4 External Control (Optional)

SNSim provides an optional external control interface for injecting exogenous signals (e.g., external events or policy changes) and for controlling agent activation schedules. In the experimental setting of this paper, External Control is used to determine which agent acts at each simulation step, so as to align the simulation timeline with the posting order observed in the realworld data. Through this mechanism, each tweet in the dataset can be paired with a corresponding simulated post, enabling direct alignment and comparison with ground truth data.

We emphasize that this control is introduced solely for experimental alignment and does not affect the autonomous decision making processes of the Profile, Cognition, or Rewire modules. In this sense, External Control is treated as an experimental alignment mechanism rather than a required component of the framework. In future work, this module can be extended to simulate different policy interventions or external shocks and to analyze their effects on the formation, diffusion, and polarization of collective opinions.

E Base Model Prompt

This section documents the prompt template used by the *Base Model* as a reference condition in our experiments. Unlike SNSim, the Base Model does not incorporate explicit memory construction, identity conditioning, or network-aware reasoning. Agents are treated as stateless text generators that react only to their immediate local context.

Specifically, the Base Model relies on a single prompt for content generation based on (i) the agent's recent posting history and (ii) the set of messages currently visible on its screen. No personal summaries, group-level abstractions, or reflective cognition are provided to the language model.

E.1 Content Generation Prompt

At each activation step, the agent observes a bounded set of surrounding messages and is instructed to generate a new post that aligns with its current opinion and the prevailing discourse. The prompt does not encode any explicit notion of social identity, memory abstraction, or long-term behavioral consistency.

Base Model Prompt for Content Generation

Here are the history tweets of your own: {history_tweets}.
Here are the surrounding tweets: {surrounding_tweets}.
Based on these surrounding tweets and your own opinion, generate a tweet that aligns with your view and reflects the current tweet trends.
Use the following format:
New Tweet: <tweet>

E.2 Stance Inference Prompt

After a new tweet is generated, the Base Model applies a separate stance inference prompt to map the generated text to a discrete opinion label. This step is used solely for evaluation and state updating, and does not feed back into future prompt construction.

The stance inference prompt follows a standard instruction-based classification format, without access to agent memory or social context beyond the generated tweet itself.

Base Model Prompt for Stance Inference

Instruction: Analyze the tweet below in the following context: [{topic}].
Consider the text, subtext, regional and cultural references, and any implicit meanings to determine the stance expressed in the tweet towards the target.
The possible stances are:

- FAVOR: The tweet has a positive or supportive opinion towards the target.
- AGAINST: The tweet opposes or criticizes the target.
- NONE: The tweet is neutral or does not express a stance.

Tweet: [{new_tweet}]
Question: What is the stance expressed in the tweet towards the target "[{target}]"?
Choose one of the following options: FAVOR, AGAINST, NONE.
Show your reason.
Answer:

Discussion. The Base Model prompt serves as a minimal reference that reflects common LLM-based social simulation practices in which agents respond to local textual input without explicit modeling of identity, memory, or social structure. By contrasting this setting with SNSim, we isolate the contribution of structured prompting and modular cognition to simulation fidelity.

F Prompt Templates Used in SNSim

This appendix documents the prompt templates used in SNSim for memory construction, content generation, stance inference, and network rewiring. Rather than treating prompts as implementation artifacts, SNSim uses prompts as explicit interfaces between symbolic simulation states and LLM-based reasoning.

Each prompt corresponds to a well-defined operation in the simulation loop (Algorithm 1), with function names explicitly matching those used in the algorithm (e.g., REFLECT, LLM_GENPOST, OPINION_CLASSIFY, REWIRE).

F.1 Memory Construction Prompts (REFLECT, UPDATEMEM)

The memory construction prompts operationalize the memory component of the Cognition module and implement the REFLECT and UPDATEMEM operations in Algorithm 1. Instead of maintaining numerical hidden states, SNSim represents memory as natural-language summaries that are injected into downstream prompts. This design reflects the assumption that agents recall and reason about past interactions through high-level narratives rather than exact message logs.

Different memory prompts correspond to different assumptions about how individuals integrate personal experience and group-level signals when forming stances.

Personal memory. The personal memory prompt summarizes an agent's own historical posts into a concise identity-consistent narrative. This summary captures stable preferences, values, and rhetorical tendencies, and serves as a personalized prior that conditions future content generation.

Prompt for Personal Memory Summarization

You are summarizing a user's personal experiences on social media, with an emphasis on how these experiences reflect their identity and role within a broader community.
Create a concise summary that captures the user's overall perspective, values, or tendencies as revealed in the posts. Focus on recurring patterns rather than individual messages.
Personal Posts: """ {history_tweet} """

Group-level memory. The group memory prompt abstracts recent posts from an agent's social neighborhood into a shared group-level narrative. This representation models how individuals infer collective identity and dominant viewpoints from their local social context, rather than tracking each neighbor independently.

Prompt for Group Memory Summarization

You are summarizing the collective experiences and viewpoints of a social group that shares a common identity.
Based on the provided posts from group members, generate a concise summary that reflects the group's shared narrative, values, or ideological tendencies.
Group Posts: """ {group_posts} """

Combined reflective memory (REFLECT). The reflection memory prompt integrates personal and group-level summaries into a single reflective narrative. This step models higher-order reasoning in which agents reconcile individual experience with perceived group identity, enabling temporally coherent stance formation. The resulting reflection is used as a compact memory representation in subsequent content generation and rewiring decisions.

Prompt for Reflection Memory Generation

You are a thoughtful social media user who reflects not only on your personal experience but also on your group identity and shared beliefs.
Given the following two inputs:
1. Personal Experience: "{personal_history}"
2. Group Perspective: "{group_history}"
Write a concise reflection (1–3 sentences) that integrates both personal and group experiences. Avoid simply restating the summaries; instead, reflect on what they imply about your values, priorities, and current stances.

F.2 Tweet Generation Prompt (LLM_GENPOST)

This prompt implements the LLM_GENPOST operation in the Cognition module. At each activation, the agent receives a bounded local context consisting of its identity-conditioned profile, optional memory representation, and a fixed window of recently visible neighbor posts.

The prompt constrains the LLM to generate a single post in a standardized format, ensuring that linguistic generation remains compatible with downstream stance inference and evaluation.

Prompt for Tweet Generation

Agent Name: {agent_name}
Role Description: {profile_description}
Historical Tweets: {history_tweets}
Memory Context: {memory_content}
Tweets Visible on Screen: {tweet_page}
Task: Based on all the information above, generate a tweet that aligns with your current view and reflects the ongoing discussion. Use the following format:
New Tweet: <tweet>

F.3 Stance Inference Prompt (OPINION_CLASSIFY)

SNSim decouples linguistic generation from stance estimation by using a separate stance inference prompt, corresponding to the OPINION_CLASSIFY operation in Algorithm 1. Rather than inferring stances directly from internal LLM states, each generated post is classified into a discrete stance label using a dedicated prompt.

This design enables consistent comparison between simulated and real-world data, even when surface linguistic realizations differ. Discrete stance labels are subsequently mapped to internal opinion scores used by the simulator.

Prompt for Stance Inference

Analyze the tweet below in the following context: [{topic}]. Determine the stance expressed toward the target [{target}].
Possible stances:
- FAVOR: supportive or positive stances.
- AGAINST: opposing or critical stances.
- NONE: neutral or no clear stance.

Tweet: "{text}"
Output only one label: FAVOR, AGAINST, or NONE.

F.4 Rewiring Decision Prompt (REWIRE)

The rewiring prompt operationalizes the REWIRE module by framing network adaptation as an identity- and content-aware decision. The LLM is provided with recent posts from current neighbors and a bounded set of candidate users, together with the agent's profile and memory representation.

The prompt requires the model to explicitly justify tie dissolution and formation, making rewiring decisions interpretable and traceable. While the experiments in this paper maintain approximately constant network density, the prompt formulation itself supports more general structural adaptation policies.

Prompt for Rewiring Decision

Agent Name: {agent_name}
Role Description: {profile_description}
Personal Experience Summary: {personal_history}
Reflection Memory: {reflection_memory}
Recent Tweets from Current Friends: {friend_tweets}
Tweets from Candidate Users: {candidate_tweets}
Task: Suggest one current friend to unfollow and one candidate user to follow, based on opinion alignment, group identity, and consistency with your stance.
Provide the output strictly in the following format:
Reason: <brief explanation> Unfollow_ID: <friend_id> Follow_ID: <candidate_id>

G More Ablation Studies

G.1 Profile Analysis

HiSim (Mou et al., 2024) adopts a heuristic profile prompting strategy, injecting relatively stable attributes inferred from user history (e.g., demographics, activity, influence, and self descriptions) into the LLM prompt. These attributes are treated as empirical *identity cues* that constrain generation style and behavior, improving controllability and heterogeneity, but without explicit theoretical grounding.

Personal retains only individual-level identity cues inferred from a user's own historical behavior, isolating the effect of *individual identity* on language and stance while excluding community-level group signals. **Compose** further incorporates community-level profiles, jointly conditioning generation on individual and group identities to provide both personal variation and group-aligned contextual constraints.

All three profiles are constructed from user historical data.

Table 4 reports the average stance error and language diversity across profile configurations. *Compose* achieves the lowest stance error, indicating that combining individual and community level identity cues substantially improves stance stability. In contrast, *Personal* and *HiSim* exhibit weaker stance performance, suggesting that individual only or heuristic profiles provide insufficient constraints on attitudinal behavior.

Regarding language diversity, *HiSim* yields the highest lexical and semantic diversity, reflecting its weakly structured profile design, while *Personal* improves syntactic diversity. *Compose* slightly reduces lexical and semantic diversity but achieves the highest syntactic diversity, indicating that group identity cues regularize semantic expression while preserving structural variation.

Overall, these results reveal a clear trade off: group-level identity conditioning enhances stance stability but constrains linguistic diversity, whereas individual-level or heuristic profiles favor expressive diversity at the expense of attitudinal consistency.

Profile	Stance	Lexical	Semantic	Syntactic
HiSim	0.6169	**0.0529**	**0.4263**	0.3763
Personal	0.6474	0.0525	0.4213	0.3878
Compose	**0.3734**	0.0411	0.3363	**0.3898**

Table 4: Average profile-level stance error and language diversity across COVID-19, Bluesky, and Ukraine datasets.

G.2 Cognition Analysis

We evaluate four cognition strategies that differ in how historical information is structured and presented to the LLM. *RawFriends* conditions generation only on recently observed neighbor tweets, without any explicit reflection or memory abstraction. *Personal* performs reflection over an agent's own historical tweets to construct personal memory. *Group* reflects over surrounding neighbors' tweets to form group-level memory. *Combine* integrates both personal and group reflections, jointly exposing the agent to self-consistent and community-level signals.

Table 5 reports the average stance error and language diversity across datasets. *Combine* achieves the lowest stance error, indicating that integrating personal and group reflections provides the strongest constraint on coherent opinion evolution.

Personal also improves stance accuracy relative to *RawFriends*, suggesting that self-reflection helps stabilize opinions by reinforcing internal consistency. In contrast, *RawFriends* exhibits the highest stance error, showing that directly reacting to recent neighbor content without abstraction introduces noise into stance updates.

In terms of language diversity, *RawFriends* achieves the highest lexical, semantic, and syntactic diversity, reflecting the unfiltered exposure to heterogeneous neighbor content. *Group* preserves relatively high diversity—particularly in semantic and syntactic dimensions—while improving stance accuracy, indicating that group-level reflection provides a stabilizing effect without fully collapsing linguistic variation. *Personal* and *Combine* yield lower diversity scores, suggesting that reflection over memory abstracts and regularizes language, trading expressive diversity for attitudinal stability.

Overall, these results highlight the central role of cognition design in balancing stability and diversity in LLM-driven agents. Explicit reflection mechanisms—especially when combining personal and group memories—are critical for producing coherent and realistic opinion dynamics.

Cognition	Stance	Lexical	Semantic	Syntactic
RawFriends	0.8049	**0.0552**	**0.4895**	**0.4195**
Personal	0.7600	0.0484	0.3938	0.3855
Group	0.7699	0.0540	0.4458	0.4137
Combine	**0.7383**	0.0484	0.4065	0.3861

Table 5: Average stance error and language diversity metrics, aggregated across COVID-19, Bluesky, and Ukraine datasets, under different cognition settings.

G.3 Network Structure

Table 6 reports the impact of Profile (P) and Cognition (C) on network structure when rewiring is enabled. Overall, rewiring dominates network evolution, while P and C act as secondary modifiers.

Modularity does not improve with the inclusion of P or C; instead, a slight increase in error is observed, especially when both modules are enabled. This suggests that global community structure is primarily governed by rewiring, with limited contribution from agent-level cognition.

In contrast, clustering coefficient and average path length benefit consistently from P and C. Profile exhibits a slightly stronger effect on clustering, indicating that identity-related attributes mainly shape local interaction patterns. Average path length shows the largest sensitivity, with either module substantially reducing error, while combining both yields diminishing returns, implying overlapping roles in shaping global connectivity.

Assortativity remains largely unchanged across all configurations, confirming that homophily-related structure is determined mainly by the rewiring mechanism itself.

Overall, these results indicate a clear division of roles: rewiring drives large-scale network structure, whereas Profile and Cognition refine local connectivity and efficiency without altering the global topology.

Table 6: Relative Improvement of M and P Mechanisms on Network Metrics Compared to Baseline (R=1).

Config.	Mod.	Clust.	PL	Assort.
C=0, P=1	-11.88%	+25.58%	+17.88%	+2.19%
C=1, P=0	-5.83%	+16.39%	+19.44%	+3.44%
C=1, P=1	-15.34%	+29.09%	+17.33%	+2.77%

H Extended Discussion of Limitations

Dependence on prompt design and LLM priors. SNSim realizes social simulation through structured prompting and LLM-driven decisions, treating prompts as explicit interfaces between symbolic simulation states and model behavior. While this design substantially improves controllability and interpretability compared to ad hoc prompting, it also means that SNSim inevitably inherits known sensitivities of large language models. In particular, LLM behavior can vary with prompt phrasing, domain context, and model-specific training distributions, leading to non-trivial differences across topics and datasets.

More fundamentally, many foundation models exhibit strong normative or safety-aligned priors learned during pretraining and alignment, such as a tendency to favor mainstream public health narratives or to adopt stereotypical positions in highly polarized political contexts. These priors may manifest as systematic ideological skew in generated content and downstream decisions (e.g., stance inference or rewiring preferences), even when agents are conditioned on identity-specific profiles. While SNSim's modular prompt design—especially identity conditioning and reflective cognition—can partially counteract such tendencies by grounding generation in individual and community context, prompt engineering alone cannot fully eliminate

biases that are internal to the model. Addressing this limitation likely requires complementary approaches such as model selection, fine-tuning, debiasing objectives, or explicit counterfactual supervision, which we leave to future work.

Opinion representation and evaluation granularity. SNSim represents opinions using discrete stance labels (FAVOR / AGAINST / NONE) and evaluates simulation fidelity by comparing stance trajectories against temporally aligned ground truth. This choice ensures compatibility with widely used stance detection datasets and enables direct, interpretable error metrics. However, such a representation inevitably abstracts away finer-grained opinion states, including ambivalence, uncertainty, conditional support, or gradual opinion drift across related topics.

In addition, stance inference errors may propagate into downstream components of the simulation. For example, misclassified posts can influence memory summarization, reflection, and rewiring decisions, thereby compounding early errors over time. Moreover, stance alignment focuses primarily on target-specific opinions and does not explicitly capture other psychologically relevant dimensions of social behavior, such as confidence, moral framing, emotional intensity, or rhetorical strategy. Future extensions of SNSim could incorporate continuous opinion scores, calibration-aware evaluation, or multi-dimensional opinion representations, as well as complementary metrics that assess temporal coherence, volatility, and semantic faithfulness beyond coarse categorical labels.

Rewiring mechanism and platform-specific dynamics. The REWIRE module models network evolution through bounded candidate selection and identity- and content-aware tie adaptation, capturing key homophily-driven tendencies observed in social networks. While this abstraction allows for interpretable and controllable network evolution, it does not explicitly model several platform-specific mechanisms that strongly shape real-world interaction dynamics. These include algorithmic recommendation systems, content ranking and exposure effects, retweet or repost cascades, moderation policies, and visibility constraints, all of which can substantially influence who interacts with whom and when.

Furthermore, in the current experimental setting, rewiring decisions are constrained to maintain approximately stable network density in order to facilitate controlled comparison with observed networks. This design choice may underrepresent regimes in which networks undergo significant expansion or contraction, or where tie changes are highly asymmetric across users (e.g., influencer-driven growth or mass unfollow events). Extending SNSim to incorporate explicit platform mechanisms and more flexible structural constraints is an important direction for future work, particularly for studying intervention effects or large-scale structural shifts.

Ethical considerations of simulating social behavior. As with other LLM-driven simulation frameworks, SNSim may reproduce or amplify harmful stereotypes, misinformation-like narratives, or biased representations of social groups, especially when modeling highly polarized or sensitive topics. Although SNSim is intended as a methodological tool for studying social dynamics rather than generating persuasive or prescriptive content, simulated outputs could still be misused if deployed without appropriate safeguards.

These risks underscore the importance of responsible use, including careful dataset selection, transparent reporting of model limitations, and explicit usage constraints when applying social simulation frameworks in sensitive contexts. Ethical considerations should be treated as an integral part of methodological design rather than an afterthought, particularly as LLM-based simulations become more realistic and easier to scale.

I Other Results

Profile	Stance	Lexical	Semantic	Syntactic
Hisim	0.6866	0.0528	0.3478	0.3783
Personal	0.7651	0.0507	**0.3679**	**0.3832**
Compose	**0.6681**	**0.0538**	0.3571	0.3790

Table 7: Profile-level stance and language metrics (smaller MSE is better; larger diversity is better).

Profile	Stance	Lexical	Semantic	Syntactic
HiSim	0.941	0.070	0.589	0.368
SIT_Personal	0.950	**0.072**	**0.570**	0.388
SIT_Compose	**0.226**	0.035	0.326	**0.395**

Table 8: Comparison of stance error and language diversity metrics on the Bluesky dataset.

Profile	Stance	Lexical	Semantic	Syntactic
HiSim	**0.2230**	**0.0359**	**0.3422**	0.3826
SIT_Personal	0.2270	0.0349	0.3259	0.3922
SIT_Compose	0.2262	0.0346	0.3258	**0.3953**

Table 9: Profile-level stance and language metrics for the Ukraine dataset (smaller MSE is better; larger diversity is better).

Cognition	Stance	Lexical	Semantic	Syntactic
RawFriends	0.7921	0.0585	**0.4077**	0.4229
Personal	**0.6971**	0.0512	0.3439	0.4055
Group	0.7366	**0.0587**	0.3880	**0.4340**
Combine	0.7966	0.0434	0.3491	0.3941

Table 10: Cognition-level stance and language metrics on the COVID-19 dataset (smaller MSE is better; larger diversity is better).

Cognition	Stance	Lexical	Semantic	Syntactic
RawFriends	0.9814	0.0697	**0.6400**	**0.4137**
Personal	0.9589	0.0595	0.5087	0.3610
Group	0.9639	0.0648	0.5383	0.3847
Combine	**0.9404**	**0.0715**	0.5623	0.3887

Table 11: Bluesky dataset: Cognition-level stance and language metrics (smaller MSE is better; larger diversity is better).

Cognition	Stance	Lexical	Semantic	Syntactic
Rawfriends	0.6412	0.0375	**0.4207**	0.4220
Personal	0.6239	0.0346	0.3289	0.3899
Group	0.6093	**0.0384**	0.4110	**0.4223**
Combine	**0.4778**	0.0304	0.3080	0.3756

Table 12: Cognition-level stance and language metrics on the Ukraine dataset (smaller Stance MSE is better; larger diversity is better).

Cognition	Stance	Lexical	Semantic	Syntactic
RawFriends	0.8049	**0.0552**	**0.4895**	**0.4195**
Personal	0.7600	0.0484	0.3938	0.3855
Group	0.7699	0.0540	0.4458	0.4137
Combine	**0.7383**	0.0484	0.4065	0.3861

Table 13: Average cognition-level stance error and language diversity across COVID-19, Bluesky, and Ukraine datasets. Smaller stance error is better; larger diversity scores indicate higher linguistic diversity.

Attribute	**Unfollowed Friend**	**Followed Candidate**
Tweet Content	*"Post viral fatigue is not unique to Covid... it's a lobby invention..."*	*"Maybe when vaccine companies become liable... until then... NO THANKS"*
Key Theme	Skepticism of medical narratives (General)	Accountability & Individual Freedom (Specific)
Alignment	High (Matches current views)	Very High (Resonates with core values)
Novelty	Low (Redundant information)	High (Reinforces specific political stance)
Decision	Unfollow	Follow

Table 14: Example of SNSim Rewire decision with friend–candidate comparison.

Case	t	Dropped neighbor (tweet gist)	Added candidate (tweet gist)	LLM rationale (compressed)
C1	297	F_{drop}: politicized conspiratorial claim (e.g., suggesting a "mask vaccine" was used in a political campaign)	C_{add}: accountability-framed skepticism (e.g., emphasizing manufacturer liability and incentives for safety)	Rejects low-credibility/off-topic content; prefers value-consistent, policy-oriented skepticism.
C2	521	F_{drop}: authority-appeal forecast (e.g., "normality only after next-generation vaccines")	C_{add}: concrete policy/logistics discussion (e.g., procurement volumes, shortage risks, and hesitancy)	Prioritizes actionable, context-rich information that supports the agent's institutional skepticism.

Table 15: Two representative LLM-driven rewiring decisions in SNSim (IDs anonymized). Each case selects one neighbor to unfollow and one candidate to follow, alongside a concise rationale grounded in content alignment and perceived usefulness/credibility.

Baseline	ΔMSE$\downarrow$	ΔMAE$\downarrow$	$\Delta R^2 \uparrow$	Lexical$\uparrow$	Semantic$\uparrow$	WL$\uparrow$
HiSim	+12.93	+5.45	+150.5	+58.2	+2.81	+29.1
SOP	+8.12	+4.19	+229.2	+319.5	+206.5	+18.0
FPS	+30.86	+23.03	+112.6	+78.5	+85.6	+13.6

Table 16: Improvement (%) of the proposed model over existing baselines on stance prediction and language diversity metrics for the COVID dataset. Network dynamics are not considered in these baselines.

Category	Metric	FPS	SOD	HiSim	Base Model	SNSim	Improvement
Language	Lexical ↑	0.033	0.014	0.037	0.042	**0.059**	40%
	Semantic ↑	0.329	0.199	0.594	0.578	**0.610**	6%
	Syntactic (WL) ↑	0.364	0.350	0.320	0.384	**0.413**	8%
Opinion	Stance MSE ↓	0.867	0.652	0.688	0.900	**0.599**	33%
	Stance MAE ↓	0.562	0.451	0.457	0.569	**0.432**	24%
	Stance R^2 ↑	-0.378	-0.037	-0.094	-0.430	**0.047**	111%

Table 17: Performance comparison on the **COVID-19 dataset**. FPS, SOD, HiSim, the Base Model, and the full SNSim framework are evaluated under identical observational inputs. Arrows (↑ / ↓) indicate whether higher or lower values correspond to better performance. Language metrics measure lexical, semantic, and syntactic diversity, where syntactic diversity is computed using the Weisfeiler–Lehman (WL) kernel. Opinion metrics evaluate stance trajectory alignment against ground truth. Network metrics assess structural realism relative to the observed network. The improvement column reports the relative change of SNSim over the Base Model.

Category	Metric	FPS	SOP	HiSim	Base Model	SNSim	Improvement
Language	Lexical ↑	0.020	0.010	0.037	0.048	**0.054**	11 %
	Semantic ↑	0.298	0.181	**0.548**	0.342	0.374	9 %
	Syntactic (WL) ↑	0.328	0.343	0.321	0.333	**0.375**	13 %
Opinion	Stance MSE ↓	0.306	0.274	0.275	0.251	**0.240**	4 %
	Stance MAE ↓	0.477	0.438	0.441	0.407	**0.389**	4 %
	Stance R^2 ↑	-2.591	-2.212	-2.225	-1.943	**-1.813**	7 %

Table 18: Overall comparison of stance alignment and language diversity on the **Ukraine war dataset**. FPS, SOP, HiSim, the Base Model, and the full SNSim framework are evaluated under identical observational inputs. Arrows (↑ / ↓) indicate whether higher or lower values correspond to better performance. Syntactic diversity is measured using the Weisfeiler–Lehman (WL) kernel. The improvement column reports the relative change of SNSim over the Base Model.

Category	Metric	FPS	SOP	HiSim	Base Model	SNSim	Improvement
Language	Lexical ↑	0.023	0.013	0.034	0.038	**0.063**	65 %
	Semantic ↑	0.414	0.318	0.548	0.515	**0.647**	26 %
	Syntactic (WL) ↑	0.332	0.355	0.259	0.384	**0.417**	9 %
Opinion	Stance MSE ↓	0.920	0.802	0.698	**0.765**	0.783	-2 %
	Stance MAE ↓	0.805	0.715	0.600	**0.642**	0.659	-3 %
	Stance R^2 ↑	-2.871	-2.375	**-1.939**	-2.223	-2.296	-3 %

Table 19: Overall comparison of stance alignment and language diversity on the **Bluesky dataset**. FPS, SOP, HiSim, the Base Model, and the full SNSim framework are evaluated under identical observational inputs. Arrows (↑ / ↓) indicate whether higher or lower values correspond to better performance. Syntactic diversity is measured using the Weisfeiler–Lehman (WL) kernel. The improvement column reports the relative change of SNSim over the Base Model.

Category	Metric	FPS	SOP	HiSim	Base Model	SNSim	Improvement
Language	Lexical ↑	0.025	0.012	0.036	0.043	**0.059**	37 %
	Semantic ↑	0.347	0.233	**0.597**	0.478	0.544	14 %
	Syntactic (WL) ↑	0.341	0.349	0.300	0.367	**0.402**	10 %
Opinion	Stance MSE ↓	0.698	0.576	0.554	0.639	**0.541**	15 %
	Stance MAE ↓	0.615	0.535	0.499	0.539	**0.493**	9 %
	Stance R^2 ↑	-1.947	-1.541	-1.419	-1.532	**-1.354**	12 %

Table 20: Average performance across three datasets (main experiment, Ukraine, and Bluesky), evaluating stance alignment and language diversity. All values are computed as arithmetic means over datasets. Network metrics are omitted for clarity.