Learning to Make Friends: Coaching LLM Agents toward Emergent Social Ties

Philipp J. Schneider
EPFL
Lausanne, Switzerland
philipp.schneider@epfl.ch

Lin Tian
University of Technology Sydney
Sydney, Australia
lin.tian-3@uts.edu.au

Marian-Andrei Rizoiu
University of Technology Sydney
Sydney, Australia
marian-andrei.rizoiu@uts.edu.au

Abstract

Can large language model (LLM) agents reproduce the complex social dynamics that characterize human online behavior—shaped by homophily, reciprocity, and social validation—and what memory and learning mechanisms enable such dynamics to emerge? We present a multi-agent LLM simulation framework in which agents repeatedly interact, evaluate one another, and adapt their behavior through in-context learning accelerated by a coaching signal. To model human social behavior, we design behavioral reward functions that capture core drivers of online engagement, including social interaction, information seeking, self-presentation, coordination, and emotional support. These rewards align agent objectives with empirically observed user motivations, enabling the study of how network structures and group formations emerge from individual decision-making. Our experiments show that coached LLM agents develop stable interaction patterns and form emergent social ties, yielding network structures that mirror properties of real online communities. By combining behavioral rewards with in-context adaptation, our framework establishes a principled testbed for investigating collective dynamics in LLM populations and reveals how artificial agents may approximate or diverge from human-like social behavior.

1 Introduction

Have you ever been convinced by an AI? A recent study on Reddit [34]—where AI-generated content was used to test how effectively agents could persuade users—sparked public backlash. Conducted without users' consent, the experiment was widely condemned for crossing ethical and legal boundaries. But why would anyone pursue such a study? And why does the scientific community remain deeply invested in modeling human behavior online?

The short answer is that social media has permeated nearly every aspect of society. It shapes how information flows, blurs the line between public and private spheres, and serves simultaneously as a news source, marketplace, and social infrastructure. This raises a fundamental question: can we, without compromising individual privacy, construct a digital twin that authentically captures the behavior of real social media users?

While the linguistic, psychological, and cognitive capabilities of individual large language model (LLM) agents have been demonstrated [6, 28, 36], their collective behavior in networked environments

remains far less understood. The implications of building such digital twins are far-reaching [4]. Policymakers could use them to test the effects of moderation strategies—such as content removal [37]—before deploying them at scale. Researchers could simulate the spread of opinions, the formation of online communities, or the contagion of consumer behavior. Security analysts could explore how to defend against manipulation or detect coordinated influence operations [2].

In this paper, we introduce a framework for multi-agent LLM conversations designed to simulate both private and public user discussions. By equipping agents with reward-based tasks, we move beyond the recently demonstrated agent-based modeling (ABM) capabilities of LLMs and leverage their cognitive abilities to study the emergence of social ties within dynamic simulations.

1.1 Unique Challenges

Although prior work has explored multi-agent LLM frameworks, simulating social media conversations raises three distinct challenges:

Agentic Personas. LLMs are trained to minimize prediction error in text generation, but this does not guarantee realistic user personas. Poorly specified personas often exhibit similar linguistic patterns, failing to reflect the diversity of real online users. Authentic simulations require agents that vary in values, norms, and cognitive styles, supported by memory structures that track interactions and evolve over time.

Learning Mechanisms. Games and other structured environments provide explicit rewards that guide agents toward effective strategies. Social media, by contrast, involves open-ended exchanges where memory must approximate the human ability to retain recent interactions while integrating longer conversational histories. Because engagement spans a spectrum—from passive awareness (viewing) to active participation (commenting, sharing)—agents need memory architectures that enable adaptive learning across contexts.

Social Topology. Users with shared interests tend to cluster, reinforcing in-group favoritism and out-group aversion that drive polarization and echo chambers. Structural features such as follower networks shape the spread of information. Yet a central open question remains: what network structures emerge when LLM agents interact at scale?

1.2 Our Contributions

To address these challenges, we introduce a multi-agent LLM platform that advances the modeling of online social dynamics through three core contributions:

Conversation Room. We build an interactive environment that integrates both public and private communication channels. This design enables agents to deploy distinct strategies across contexts and allows systematic analysis of how channel choice shapes conversational dynamics.

Reward Structures. We formalize reward functions that translate empirically grounded human motivations—including self-presentation, social interaction, coordination, and emotional support—into agent objectives. This framework establishes a principled bridge between behavioral theory and multi-agent LLM learning.

Tie Formation. We design mechanisms that allow social ties to emerge endogenously from conversational interactions, without relying on pre-defined network structures. This enables systematic study of how support, alignment, and homophily drive the formation of group structures.

2 Preliminaries

Multi-agent LLMs. Early work established that LLM agents can display human-like routines when equipped with memory and cognitive mechanisms. Park et al. [35] introduced the *observation–planning–reflection* cycle, showing that agents embedded in interactive environments develop coherent daily behaviors. Subsequent studies demonstrated that LLM populations can approximate human-level attitudes and sociodemographic effects, reproducing patterns found in surveys and experiments [1, 3]. Beyond reproducing individual or population-level behavior, frameworks such as CAMEL [25] and orchestration stacks like LangChain and AutoGen [41] enable coordination, specialization, and standardized multi-agent dialogues. Recent surveys synthesize these design patterns and emphasize the need for systematic evaluation against human baselines [18, 14].

Personas and diversity. A central challenge in simulation is ensuring credible heterogeneity across agents. Simple prompting often yields homogeneous responses, whereas personality induction

methods, such as Big Five personality prompting, can generate stable stylistic variation [21]. The PERSONA benchmark extends this approach, offering pluralistic alignment across diverse profiles [8]. At the same time, persona prompts may inadvertently surface reasoning biases [19], underscoring the importance of careful evaluation. More recent approaches move beyond prompting to train agents in simulated societies, allowing them to internalize social norms and polarization dynamics through interaction [26]. Together, these studies highlight both the promise and pitfalls of persona diversification for multi-agent LLM simulations.

From interactions to social structure. A key question is whether agent collectives reproduce network-level regularities. Emerging studies have examined LLM-generated networks, link formation, and paradoxes of visibility such as the friendship paradox [9, 33]. These experiments connect directly to canonical findings in human networks, including assortative mixing [30], modular community structure [13], and diffusion dynamics such as emotional contagion, misinformation spread, and echo chambers [24, 39, 10]. The ability of LLMs to produce persuasive or deceptive content further raises the stakes for safety-aware evaluation [44].

Our focus. Building on these streams of research, our framework centers on sequential, socially motivated behavior—spanning interaction, information seeking, self-presentation, coordination, and emotional support—operationalized through rewards, in-context coaching, and memory. We examine how such behavior gives rise to social ties, which form through micro-level signals (approval, reciprocity, latency) and aggregate into macro-level network structures (clustering, modularity, tie persistence). By allowing ties to emerge endogenously, without reliance on a fixed graph, we provide a testbed for studying phenomena such as homophily, community formation, and echo chambers.

3 Methodology

Figure 1 provides an overview of our multi-agent LLM platform for simulating social media conversations. The framework is organized into two main components: persona creation (A) and the social media simulation process (B). In the sections that follow, we describe how agents are initialized, outline the progression of the simulation, present the core of our learning paradigm—the task-specific reward structures of individual agents—and demonstrate how network structures emerge.
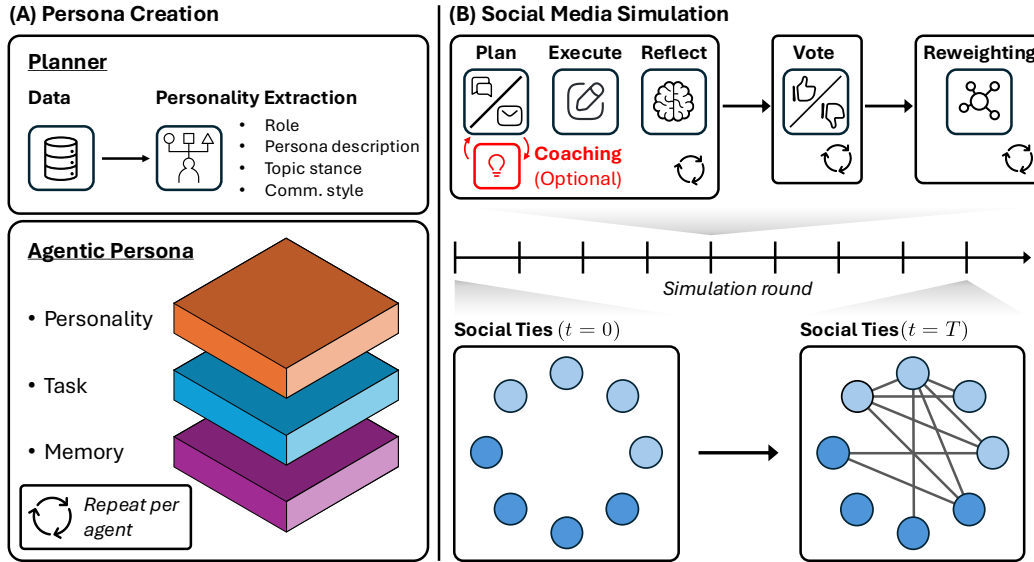

Figure 1: Multi-agent LLM social media conversation framework.

3.1 Persona Creation

Building agent-personas requires the design of psychologically grounded and behaviorally coherent profiles. To avoid randomly sampling attributes from a population of observed users, we begin by instantiating a *planner* agent. This agent orchestrates the creation of inherited persona data, drawing

on corpora extracted from real-world online discussions on a specific topic. With the growing prevalence of platforms where users remain largely anonymous, many features commonly leveraged in frameworks such as [8]—including age, sex, race, or education—are no longer observable, even though people in real interactions might infer them from user profiles. Instead, we focus on features that can be reliably inferred from the content itself: *role*, *persona description*, *topic stance*, and *communication style*. These attributes are directly grounded in the observed communications within the selected dataset.

Building on these content-based inferences, we construct agentic personas through a three-layer structure detailed below:

Personality. The first layer captures personality features derived from text data. We draw on the Big Five traits and their cross-cultural stability [11, 27], complemented by more granular facets (e.g., sociability as an Extraversion facet; stubbornness as an Agreeableness facet) and constructs beyond the Five (e.g., risk preference; interpersonal openness). High Extraversion is reflected in linguistic markers such as frequent emoji use and elevated conversational activity, whereas low Agreeableness is associated with increased negation and critical language. To assess these dimensions, we administer the Mini-IPIP, a concise Big Five inventory [12, 15].

Task. Why individuals engage on social media platforms is shaped by a range of underlying motives. Whiting and Williams [40] draw on gratification theory to explain such engagement, identifying key drivers such as social interaction, information seeking, and entertainment. Assigning a persona a specific task based on these motivations results in markedly different behaviors that often align with the platform's social engagement hierarchy—ranging from passive users (often referred to as lurkers) to highly active content creators. Related social media phenomena, such as agents exhibiting stubbornness and resistance to opinion change, can likewise be integrated into the simulation. For clarification of specific task formulations, we refer to Sec. 3.3, where we define the reward structures.

Memory. The role of memory in agents is essential for capturing the dynamics of evolving discussions. We structure persona memory across three components: *conversation memory*, which records all prior interactions; *relationship memory*, which tracks information about other agents the persona has interacted with; and *opinion memory*, which abstracts the content of a conversation into a singular entry representing a stance or belief. Together, these form a lightweight analog to long-term human memory (i.e. [43]) and enable agents to exhibit path-dependent behavior in extended dialogues.

3.2 Conversation Simulation

Once the personas are initialized and their distinct personalities verified through a pre-survey assessment, we initiate the simulation. Each agent engages in a conversation on a predefined topic. At time $t = 0$, we observe the social network $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0, A_0)$ with $\mathcal{E}_0 = \emptyset$, indicating that there are no edges and no prior knowledge among users in $\mathcal{V}$; the adjacency matrix satisfies $[A_0]_{uv} = 0$ for all $u, v \in \mathcal{V}$. Our objective is to characterize the network structure $\mathcal{G}_T$ that emerges after the agents' interactions. Messaging occurs through both direct (user-to-user) and public channels (visible to all agents), consistent with prior work showing that channel choice depends on privacy concerns, audience size, and message sensitivity. By incorporating both modes, we allow agents to adapt communication strategies to their emotional grounding and negotiation context.

In the opening round ($t = 1$), agents are unaware of one another and therefore required to make a public post (POST). In subsequent rounds, they may choose to post (POST), comment (COM), send a direct message (DM), or take no action (NOT). The per-round procedure unfolds in three phases.

Plan–Execute–Reflect. Each agent u plans a pre-specified number of actions N based on the batch of content from the previous round $\mathcal{B}_{t-1}$[1], the reward $R_{t-1}(u)$, and additional side information (see Sec. A). Actions are selected from $\mathcal{A} = \{\text{POST}, \text{COM}, \text{DM}, \text{NOT}\}$. Planning is non-trivial: each action requires further specification (e.g., which post to comment on or which user to message) in addition to the strategic objective of maximizing future reward. As an optional step, we test whether a "coach" can simplify this process by providing a tip, enabling the agent to focus on constructing a well-structured response. Once a valid plan is constructed, the agent executes each action, generating content consistent with the action type (e.g., a DM reply is based on the specific message received in the prior round). Executed actions are stored in the agent's memory, and this procedure is repeated for all agents.

Vote. Social validation is central to how individuals perceive support or friendship online. Accordingly, for all publicly visible content (POST, COM), we enable agents to cast votes by liking, disliking,

---

[1]For each user u, this includes all public content plus direct messages addressed to u.

or remaining neutral toward the content.

Reweighting. After each round, agents update tie strengths by reweighting their relationships. This process leverages multiple behavioral signals to inform decisions about whether and how to adjust connections.

3.3 Rewards for In-Context Learning

In the absence of explicit reward signals, agent behaviors tend to collapse toward purely greedy actions, failing to capture the diversity and goal-directed reactivity characteristic of real-world social systems. To address this limitation, we define reward functions that align agent learning with task objectives reflecting user motivations.

Social Interaction (SOC). Interacting with others and maintaining social ties are primary drivers of user engagement on social media platforms. To formalize this, we define the set of *direct-exchange actions* as $\mathcal{A}^{\mathtt{DIR}} \subset \mathcal{A}$, $\mathcal{A}^{\mathtt{DIR}} = \{\mathtt{DM}, \mathtt{COM}\}$, corresponding to direct messages and comments. Let $\mathcal{B}_t$ be the set of all actions taken by users in the network $\mathcal{V}$ in round t, and let $\mathcal{B}_t^{\mathtt{DIR}} = \{a \in \mathcal{B}_t : \mathrm{TYPE}(a) \in \mathcal{A}^{\mathtt{DIR}}\}$ denote the subset of actions in round t that involve direct exchanges. We define the social interaction reward $R_t^{\mathtt{SOC}} : \mathcal{V} \to [0, 1]$ as

$$R_t^{\mathtt{SOC}}(u) = (1 - \beta^{\mathtt{SOC}})\frac{I_t^{\mathtt{sent}}(u)}{N} + \beta^{\mathtt{SOC}}\frac{I_t^{\mathtt{rec}}(u)}{|\mathcal{B}_t^{\mathtt{DIR}}|},$$

where $I_t^{\mathtt{sent}}(u) := \sum_{a \in \mathcal{B}_t^{\mathtt{DIR}}} \mathbb{1}\{a : u \to \cdot\}$, and $I_t^{\mathtt{rec}}(u) := \sum_{a \in \mathcal{B}_t^{\mathtt{DIR}}} \mathbb{1}\{a : \cdot \to u\}$, with $\beta^{\mathtt{SOC}} \in [0, 1]$. Here, $I_t^{\mathtt{sent}}(u)$ counts the number of direct exchanges *sent* by user u in round t (i.e., actions with sender u and any recipient, denoted $u \to \cdot$), while $I_t^{\mathtt{rec}}(u)$ counts the number of direct exchanges *received* by user u. Both terms are normalized to ensure $R_t^{\mathtt{SOC}}(u) \in [0, 1]$, where $N = |\mathcal{B}_t(u)|$ denotes the constant number of actions by any user $u \in \mathcal{V}$ per round. The hyperparameter $\beta^{\mathtt{SOC}}$ trades off receiving and sending.

Information-Seeking Reward (INF). Users on social platforms often seek to discover new topics while maintaining a diverse information consumption. To formalize this behavior, we assume that each content item c_t^u presented to user u at time t carries a single *topic label* $\tau(c) \in \{1, 2, \ldots, K^{(\tau)}\}$, where $K^{(\tau)}$ denotes the total number of topics available on the platform. Let $\mathcal{C}_t(u)$ denote the set of items recommended to user u in round t, and define the set of topic labels encountered by u up to round $t-1$ as $\mathcal{H}_{t-1}^\tau(u) = \{\tau(c) : c \in \mathcal{H}_{t-1}(u)\}$. We define *topic novelty* as the number of previously unseen topics in the current recommendation $\mathcal{T}_t(u) = |\{\tau(c) : c \in \mathcal{C}_t(u)\} \setminus \mathcal{H}_{t-1}^\tau(u)|$, capturing the user's opportunity to encounter new information. In addition, we measure *topic diversity* using the empirical entropy over the topic distribution in $\mathcal{C}_t(u)$. Let $p_{t,k}(u) := |\{c \in \mathcal{C}_t(u) : \tau(c) = k\}|/|\mathcal{C}_t(u)|$ for $k = 1, \ldots, K^{(\tau)}$ denote the empirical frequency of topic k. The corresponding Shannon entropy is $H_t(u) := -\sum_{k=1}^{K^{(\tau)}} p_{t,k}(u)\log p_{t,k}(u)$, which satisfies $0 \le H_t(u) \le \log K^{(\tau)}$. The per-round, per-user information-seeking reward is then defined as

$$R_t^{\mathtt{INF}}(u) = (1 - \beta^{\mathtt{INF}})\frac{\mathcal{T}_t(u)}{K^{(\tau)}} + \beta^{\mathtt{INF}}\frac{H_t(u)}{\log K^{(\tau)}},$$

where $\beta^{\mathtt{INF}} \in [0, 1]$ controls the trade-off between novelty and diversity. This ensures $R_t^{\mathtt{INF}}(u) \in [0, 1]$, where $\beta^{\mathtt{INF}} = 0$ prioritizes the discovery of new topics, while $\beta^{\mathtt{INF}} = 1$ encourages consuming a broad mix of content within the current topic space.

Self-Presentation Reward (PRE). Self-presentation is a fundamental motive for user participation on social media, where individuals share content to express identity, gain visibility, and seek validation through social feedback. To formalize this, we consider that in each round t, agent u executes a batch of actions $\mathcal{B}_t(u) \subset \mathcal{B}_t$, $\mathcal{B}_t(u) = \{a_{t,1}(u), \ldots, a_{t,N}(u)\}$, where each action belongs to the action set $\mathcal{A}$. We define the subset of *self-presentation actions* as $\mathcal{B}_t^{\mathtt{POST}}(u) = \{a \in \mathcal{B}_t(u) : \mathrm{TYPE}(a) = \mathtt{POST}\}$, representing all posts created by user u in round t. The cardinality $|\mathcal{B}_t^{\mathtt{POST}}(u)|$ measures the quantity of self-generated content.

Beyond the act of posting, users also derive utility from *feedback* received on their content, reflecting community approval or disapproval. For each post $a \in \mathcal{B}_t^{\mathtt{POST}}(u)$, let $\ell_t^+(a)$ and $\ell_t^-(a)$ denote the number of likes and dislikes received, respectively. The total positive and negative feedback accumulated by v on their posts during round t are then given by

$$L_t^+\left(\mathcal{B}_t^{\mathtt{POST}}(u)\right) = \sum_{a \in \mathcal{B}_t^{\mathtt{POST}}(u)} \ell_t^+(a), \quad L_t^-\left(\mathcal{B}_t^{\mathtt{POST}}(u)\right) = \sum_{a \in \mathcal{B}_t^{\mathtt{POST}}(u)} \ell_t^-(a).$$

The self-presentation reward combines the incentive to post content with the desire to receive positive social feedback. Formally, we define

$$R_t^{\text{PRE}}(u) = (1 - \beta^{\text{PRE}})\frac{|\mathcal{B}_t^{\text{POST}}(u)|}{|\mathcal{B}_t(u)|} + \beta^{\text{PRE}}\frac{L_t^+\left(\mathcal{B}_t^{\text{POST}}(u)\right) - L_t^-\left(\mathcal{B}_t^{\text{POST}}(u)\right)}{(|\mathcal{V}| - 1)N},$$

where $\beta^{\text{PRE}} \in [0, 1]$ controls the trade-off between the intrinsic motivation to share content and the extrinsic motivation to receive social approval. The first term captures the relative frequency of self-presentation actions, while the second term measures net community feedback, normalized by the maximum possible number of other agents and actions per round to ensure scale consistency across simulations.

Coordination Reward (COORD). Coordination captures user behaviors that foster direct interaction and reciprocity within a social network, such as responding to others and explicitly mentioning peers to engage them in discussions. This reward models two forms of coordination: *visibility through mentions* and *reciprocity through replies*. We first define *mentions* as explicit references to other users (e.g., using "@user") in posts or comments. Let $\mathcal{B}_t^{\text{COORD}} = \{a \in \mathcal{B}_t : \text{TYPE}(a) \in \{\text{COM}, \text{POST}\}\}$ denote the set of all comments and posts created during round t. We define the number of times user v is mentioned in round t as $U_t^@(u) = |\{a \in \mathcal{B}_t^{\text{COORD}} : u \in \text{MENTIONS}(a)\}|$, where MENTIONS is a helper function that extracts the set of users explicitly mentioned in action a. The second form of coordination involves *reciprocity* through direct messaging. We identify the set of users who sent a direct message (DM) to user v in the previous round,

$$U_{t-1}(u) = \{v : \exists a \in \mathcal{B}_{t-1}, \text{ SNDR}(a) = v, \text{ RCPT}(a) = u, \text{ TYPE}(a) = \text{DM}\},$$

where SNDR and RCPT return the sender and recipient of action a, respectively. We then define the number of direct replies user v sends back to those who messaged them in the previous round as

$$U_t^{\text{DM}}(u) = \left|\{a \in \mathcal{B}_t : \text{TYPE}(a) = \text{DM}, \text{ SNDR}(a) = u, \text{ RCPT}(a) \in U_{t-1}(u)\}\right|.$$

This quantity captures the extent to which v reciprocates by responding to direct messages. The coordination reward is then defined as

$$R_t^{\text{COORD}}(u) = (1 - \beta^{\text{COORD}})\frac{U_t^@(u)}{|\mathcal{B}_t^{\text{COORD}}|} + \beta^{\text{COORD}}\frac{U_t^{\text{DM}}(u)}{|U_{t-1}(u)|},$$

where $\beta^{\text{COORD}} \in [0, 1]$ controls the trade-off between prioritizing *visibility* through mentions and *reciprocity* through direct replies. This formulation incentivizes agents to balance being seen by others and actively maintaining responsive communication within the network.

Emotional Support Reward (EMO). Emotional support is a core element of social interactions, reflecting the encouragement, affirmation, or criticism that users receive during their engagements. We define $\mathcal{B}_t^{\text{DIR}}$ as the set of all direct messages (DM) and comments (COM) exchanged in round t. For each action $a \in \mathcal{B}_t^{\text{DIR}}$, let $\text{RCPT}(a)$ denote the recipient and let $s(a) \in [-1, 1]$ be the sentiment score, where positive values indicate supportive content and negative values indicate hostility. We compute the total positive and negative sentiment directed toward user v in round t as

$$U_t^+(u) = \sum_{\substack{a \in \mathcal{B}_t^{\text{DIR}} \\ \text{RCPT}(a)=u}} \max\{0, s(a)\}, \quad U_t^-(u) = \sum_{\substack{a \in \mathcal{B}_t^{\text{DIR}} \\ \text{RCPT}(a)=u}} \max\{0, -s(a)\}.$$

To capture the role of sentiment in shaping user experience, we define the emotional support reward using a trade-off parameter $\beta^{\text{EMO}} \in \mathbb{R}$ and a small constant $\varepsilon > 0$ for numerical stability:

$$R_t^{\text{EMO}}(u) = \frac{1}{2}\left(1 + \frac{U_t^+(u) + \beta^{\text{EMO}}U_t^-(u)}{U_t^+(u) + |\beta^{\text{EMO}}|U_t^-(u) + \varepsilon}\right).$$

This formulation ensures that $R_t^{\text{EMO}}(u)$ remains within the unit interval. The parameter β^{EMO} allows the model to reflect different platform dynamics: positive values imply that negative sentiment contributes to engagement, negative values penalize negative sentiment more strongly, and setting $\beta^{\text{EMO}} = 1$ treats positive and negative sentiment symmetrically. Under this formulation, interactions dominated by a single sentiment polarity yield rewards close to one, while balanced sentiment or the absence of interactions result in a value near one half, indicating a neutral emotional environment.

Compositional Reward. User motivations on social platforms rarely stem from a single objective

but instead reflect a blend of *social interaction* (SOC), *information seeking* (INF), *self-presentation* (PRE), *coordination* (COORD), and *emotional support* (EMO). To capture this complexity, we define the overall reward as a weighted combination of these components, allowing the simulation to reflect diverse user goals and the interplay among engagement drivers. Formally, we write the compositional reward as

$$R_t(u) = \sum_{r \in \mathcal{R}} \lambda_r \, R_t^r(u), \quad \sum_{r \in \mathcal{R}} \lambda_r = 1, \quad \lambda_r \geq 0,$$

where the set $\mathcal{R} = \{\mathsf{SOC}, \mathsf{INF}, \mathsf{PRE}, \mathsf{COORD}, \mathsf{EMO}\}$ denotes the included reward components. The coefficients λ_r specify the relative importance of each component, enabling the model to represent users with distinct goals or to reflect platform designs that emphasize particular forms of engagement. This compositional framework ensures that while each reward remains interpretable and bounded, their aggregation can flexibly capture the multidimensional objectives underlying behavior in online environments.

3.4 Mechanisms of Social Tie Formation

We model the evolution of directed social ties over time by representing the interaction structure at each round t as a weighted, directed network $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, A_t)$, where $\mathrm{diag}(A_t) = 0$. The entry $[A_t]_{uv} \in [0, 1]$ stores the strength of the directed tie from user v to user u. Ties are initialized at zero and strengthened or weakened over time in response to interaction, allowing us to separate the formation of a tie from its subsequent development. This approach builds on actor-oriented models for dynamic social networks [38] and is consistent with empirical evidence that the persistence of communication ties depends strongly on the frequency and recency of interaction [32, 23].

To determine whether a directed tie should be updated in a given round, we introduce a binary activation denoted as $\zeta_t(v \rightarrow u) \in \{0, 1\}$ that encodes whether user v actively interacted with user u in round t. We operationalize this by defining two directed interaction channels. The ADDRESS channel activates when v directly addresses u—either by sending a private message or by publicly mentioning u. The ENGAGE channel activates when v interacts with content created by u—either through commenting or voting. Formally, let $\mathcal{B}_t$ be the batch of all actions in round t, then

$$\mathrm{ADDRESS}_{v \rightarrow u}(t) := \mathbb{1}\{\exists\, a \in \mathcal{B}_t : \; \mathsf{isDmTo}(a, u, v) \vee \mathsf{isMention}(a, u, v)\},$$
$$\mathrm{ENGAGE}_{v \rightarrow u}(t) := \mathbb{1}\{\exists\, a \in \mathcal{B}_t : \; \mathsf{isComTo}(a, u, v) \vee \mathsf{isVoteFor}(a, u, v)\},$$

and we define the overall activation as $\zeta_t(v \rightarrow u) := \mathrm{ADDRESS}_{v \rightarrow u}(t) \vee \mathrm{ENGAGE}_{v \rightarrow u}(t)$. Once activation is determined, the tie strength is updated using a gated update rule that differentiates between active and passive rounds. On active rounds, the tie is strengthened based on a scalar evidence score $e_t(v \rightarrow u) \in [0, 1]$, which aggregates the quality of the interaction. On passive rounds—when no directed interaction from v to u occurs—the tie strength decays. The update is defined as

$$[A_{t+1}]_{uv} := \begin{cases} [A_t]_{uv} + \min\left\{\Delta_{\max}, \, (1 - [A_t]_{uv})[e_t(v \rightarrow u) - \xi]_+\right\}, & \text{if } \zeta_t(v \rightarrow u) = 1, \\ (1 - \delta)[A_t]_{uv}, & \text{if } \zeta_t(v \rightarrow u) = 0, \end{cases} \quad (1)$$

where $\xi \in [0, 1)$ is a minimum-evidence threshold, $\Delta_{\max} \geq 0$ caps the per-round increase, $\delta \in [0, 1)$ controls decay, and $[x]_+ = \max\{0, x\}$. The scaling term $1 - [A_t]_{uv}$ ensures that increases become smaller as the tie approaches its maximum value, while decay acts multiplicatively to gradually fade inactive ties. To interpret δ, one can parameterize it via a half-life h such that $\delta = 1 - 2^{-1/h}$.

The evidence score $e_t(v \rightarrow u)$ is computed by aggregating several dyadic signals that capture distinct social mechanisms. These include: the *novelty* of the interaction—capturing whether v introduces new information to u, drawing on theories of brokerage and weak ties [17, 7]; the *reciprocity* of interactions [16, 31]—captured by the symmetry of feedback between users—and the *approval* expressed through likes and other positive evaluations [29]; and the *affective tone* of communication— measured by the sentiment of direct messages, reflecting relational support [20]. Each signal is bounded and passed through a monotone mapping G_θ, which ensures that stronger combinations of signals yield higher evidence scores. Further details on the construction of individual signals and the specification of G_θ are provided in the Supplementary Materials.

An alternative approach to constructing signals heuristically is to score the text from interactions in each round, as demonstrated in the prompt in Appendix A. In this approach, we replace the evidence score $e_t(v \rightarrow u)$ with a normalized score obtained from the prompt to update the ties.

For reporting purposes, we optionally export the final adjacency matrix in an undirected form, defined as $[A_t^{\mathrm{ud}}]_{uv} = \frac{1}{2}([A_t]_{uv} + [A_t]_{vu})$, which symmetrizes tie strength for visualization and summary statistics. After the final simulation round T, we apply a threshold θ to A_T^{ud} to obtain a binarized undirected graph, which is used in the network analysis. This transformation is used only for reporting and does not affect the directed tie dynamics or update rules described above.

4 Results

We propose an overall framework for studying the formation of social ties. Our experiments were conducted with $|\mathcal{V}| = 30$ agents over $T = 15$ simulation rounds and $N = 3$ actions per agent per round, with discussions centered on climate change. The persona planner utilized data from [22].[2] We present our findings, beginning with an evaluation of whether specific tasks, referred to here as policies, can be learned.
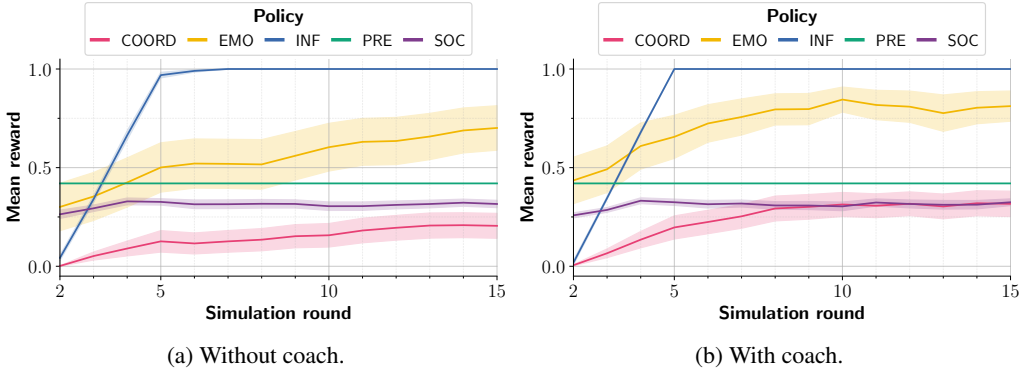

(a) Without coach.   (b) With coach.

Figure 2: Comparison between augmenting in-context learning with coach.

In Fig. 2, we observe that performance generally increases over simulation rounds, both with and without the coach. The information-seeking (INF) policy is typically the easiest to learn, as it is primarily driven by exposure to new content. Emotional support (EMO) attains comparatively high values, reflecting the generally positive behavior of most agents. Self-presentation (PRE) can be partially controlled through personal posts, whereas policies that depend on coordination (SOC, COORD) are more difficult to learn. Comparing Fig. 2a and Fig. 2b, the coach accelerates early learning for some policies but does not yield a uniform late-round improvement, with the exception of COORD and EMO, which likely benefit from improved targeting. Final performance levels are broadly similar within the variability bands. Coaching provides additional guidance that appears to reduce dispersion and structure planning, but the aggregate gains remain modest given the task complexity.

Next, we examine the reweighting step when it is performed via (i) a heuristic based on predefined signals versus (ii) a LLM text-based approach. After constructing the directed weighted graph at the final simulation step $T = 15$, we convert it into an undirected unweighted graph. The conversion threshold θ plays a central role in this process. We also allow ties to decay, and bound tie strength updates at each step by $\Delta_{\max}$. The main evaluation metrics are outlined in Appendix C. Together with baseline statistics from real networks reported in [9], they serve as external benchmarks. Comparing the two reweighting methods, we find that the heuristic approach exhibits greater variability across thresholds—particularly in density, average clustering, and average shortest path length (cf. Fig. 3). By contrast, the LLM-based text approach produces more stable results and more often falls within or near the real-network ranges for these metrics (cf. Fig. 4). Differences are less pronounced for the largest connected component, where both methods yield near-complete networks, and more modest for modularity, where both remain below real-network levels on average. We emphasize that this analysis is a first step. Further work is required to characterize how these structures evolve over time, given the dynamic interplay between agent relationships, tie-strength updates, and topical alignment.

---

[2]Due to API rate limits, the experiments presented in this work were conducted using OpenAI's GPT-4o mini.
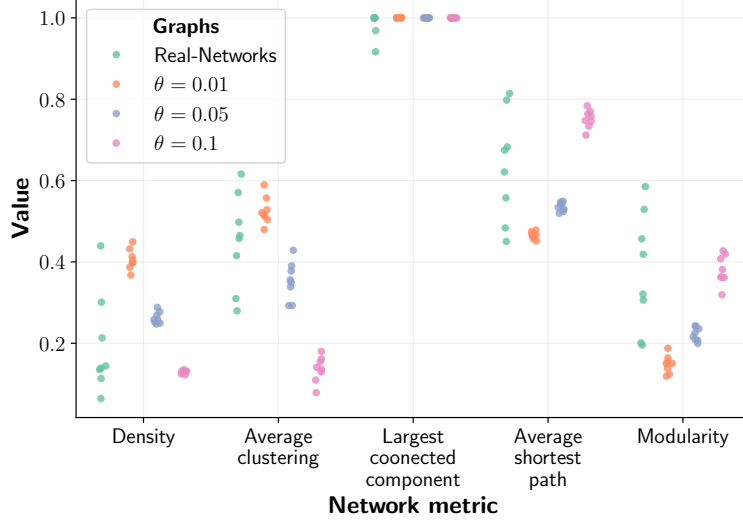
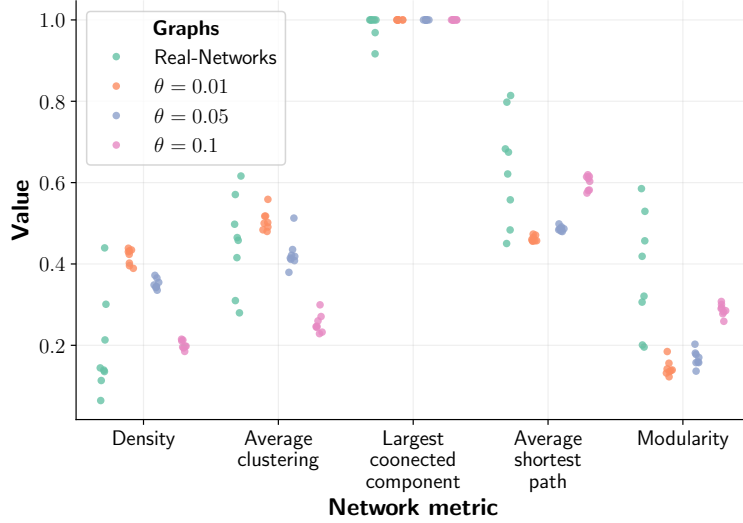Figure 3: Network metrics for heuristic tie formation.

Figure 4: Network metrics for text-based tie formation.

5 Conclusion

We present a platform that endogenously learns social ties among interacting LLM agents from their private and public exchanges. Each agent optimizes a task-specific behavioral reward grounded in gratifications theory [40] and adapts via a plan-execute-reflect loop with bandit-style updates. We show that these rewards are learnable in practice, with varying difficulty under bounded rationality and strategic influence, and that some objectives are unattainable in principle because achieving them would require control over other agents. Using the evaluation protocol of [9], the emergent networks match key statistics of real social graphs, establishing the platform as a controlled testbed for studying echo-chamber formation, the dynamics of niche communities, and the design of mitigating interventions. The study is conservative in scale ($|\mathcal{V}| = 30$, $T = 15$), uses limited replications, and starts from empty networks; scaling to larger cohorts and horizons, seeding pre-existing ties, and running intervention stress tests are clear next steps to strengthen external validity and policy relevance. Finally, with a small number of actions per round, the coaching component led to only modest gains, underscoring the difficulty of faithfully mimicking real user behavior.

References

[1] Gati V Aher, Rosa I Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *International Conference on Machine Learning*, pages 337–371. PMLR, 2023.

[2] Meysam Alizadeh, Jacob N Shapiro, Cody Buntain, and Joshua A Tucker. Content-based features predict social media influence operations. *Science Advances*, 6(30):eabb5824, 2020.

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

[4] Christopher A Bail. Can Generative AI improve social science? *Proceedings of the National Academy of Sciences*, 121(21):e2314021121, 2024.

[5] Albert-László Barabási. *Network Science*. Cambridge University Press, Cambridge, 2016.

[6] Marcel Binz and Eric Schulz. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6):e2218523120, 2023.

[7] Ronald S Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399, 2004.

[8] Louis Castricato, Nathan Lile, Rafael Rafailov, Jan-Philipp Fränken, and Chelsea Finn. PERSONA: A reproducible testbed for pluralistic alignment. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 11348–11368, 2025.

[9] Serina Chang, Alicja Chaszczewicz, Emma Wang, Maya Josifovska, Emma Pierson, and Jure Leskovec. LLMs generate structurally realistic social networks but overestimate political homophily. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pages 341–371, 2025.

[10] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118, 2021.

[11] John M Digman. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1):417–440, 1990.

[12] M Brent Donnellan, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2):192, 2006.

[13] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[14] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *Humanities and Social Sciences Communications*, 11(1):1–24, 2024.

[15] Lewis R Goldberg. The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1):26, 1992.

[16] Alvin W Gouldner. The norm of reciprocity: A preliminary statement. *American Sociological Review*, pages 161–178, 1960.

[17] Mark S Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[18] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large Language Model Based Multi-agents: A Survey of Progress and Challenges. In *International Joint Conference on Artificial Intelligence*, pages 8048–8057, 2024.

[19] Shashank Gupta, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot. Bias Runs Deep: Implicit Reasoning Biases in Persona-Assigned LLMs. In *International Conference on Learning Representations*, 2024.

[20] James S House, Karl R Landis, and Debra Umberson. Social relationships and health. *Science*, 241(4865):540–545, 1988.

[21] Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems*, 36:10622–10643, 2023.

[22] Quyu Kong, Emily Booth, Francesco Bailo, Amelia Johns, and Marian-Andrei Rizoiu. Slipping to the extreme: A mixed method to explain how extreme opinions infiltrate online discussions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 524–535, 2022.

[23] Gueorgi Kossinets and Duncan J Watts. Empirical analysis of an evolving social network. *Science*, 311(5757):88–90, 2006.

[24] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[25] Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. CAMEL: Communicative agents for "mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008, 2023.

[26] Ruibo Liu, Ruixin Yang, Chenyan Jia, Ge Zhang, Diyi Yang, and Soroush Vosoughi. Training socially aligned language models on simulated social interactions. In *International Conference on Learning Representations*, 2024.

[27] Robert R McCrae and Paul T Costa Jr. Personality trait structure as a human universal. *American Psychologist*, 52(5):509, 1997.

[28] Qiaozhu Mei, Yutong Xie, Walter Yuan, and Matthew O Jackson. A Turing test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, 121(9):e2313925121, 2024.

[29] Lev Muchnik, Sinan Aral, and Sean J Taylor. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651, 2013.

[30] Mark EJ Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[31] Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437(7063):1291–1298, 2005.

[32] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[33] Gian Marco Orlando, Valerio La Gatta, Diego Russo, and Vincenzo Moscato. Can generative agent-based modeling replicate the friendship paradox in social media simulations? In *Proceedings of the 17th ACM Web Science Conference 2025*, pages 510–515, 2025.

[34] Cathleen O'Grady. 'Unethical' AI research on Reddit under fire. *Science*, 2025.

[35] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023.

[36] Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. On the conversational persuasiveness of GPT-4. *Nature Human Behaviour*, 2025.

[37] Philipp J Schneider and Marian-Andrei Rizoiu. The effectiveness of moderating harmful online content. *Proceedings of the National Academy of Sciences*, 120(34):e2307360120, 2023.

[38] Tom AB Snijders, Gerhard G Van de Bunt, and Christian EG Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010.

[39] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[40] Anita Whiting and David Williams. Why people use social media: A uses and gratifications approach. *Qualitative Market Research*, 16(4):362–369, 2013.

[41] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. AutoGen: Enabling next-gen LLM applications via multi-agent conversations. In *Conference on Language Modeling*, 2024.

[42] Lanqin Yuan, Philipp J Schneider, and Marian-Andrei Rizoiu. Behavioral homophily in social media via inverse reinforcement learning: A Reddit case study. In *Proceedings of the ACM on Web Conference 2025*, pages 576–589, 2025.

[43] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731, 2024.

[44] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. Synthetic lies: Understanding ai-generated misinformation and evaluating algorithmic and human solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2023.

A Prompt Specifications

In this section, we present the prompts underlying our multi-agent LLM simulation: *Coach*, *Plan*, *Vote*, and *Tie-Update*. As detailed in the methodology, the *Coach* prompt provides action recommendations that streamline the planning of N actions in each simulation round t. Although the *Plan* prompt contains nearly all required information on its own, planning remains nontrivial—particularly when integrating private (direct messages) and public (posts and comments) channels—and can overload the agent. Producing a valid, executable action list is therefore essential.

The *Plan* prompt can run with or without the coach's *tip*. If a candidate plan fails our validation checks, we re-prompt (up to three times) until a valid plan is produced.

Coach Prompt

System: You are a strategy coach in a multi-round social media game. Your job is to help the agent maximize long-term rewards by closing the gaps between their current reward scores and their target weights. Priorities are ordered by importance, so the first items need the most improvement. Output 3-5 one-sentence bullet tips that are concrete, executable, and encourage variety in partners and topics across rounds to promote diversity and exploration. No preamble, no explanations. Remember: the total reward per round is capped at 1, and many reward components depend on how others engage with the agent, so your tips should encourage responses and collaboration.

User: Priority this round (high→low): {focus}
Guidelines: {guidelines}

Context:
· Users: {users}
· DM senders last round: {dm_last_senders}
· Recent partners: {recent_partners}
· Known topics so far: {seen_topics}
· Recent topics (last 3): {last_recent}

Each tip must include:
· action type (POST/COM/DM),
· target user (from lists) or 'top-liked post',
· topic (choose a relevant or new topic; exploring new topics helps INF),
· whether you invite them to @mention you back (only if aiming at COORD),
· tone (supportive/neutral/critical).

Focus items are ordered by how far below target the corresponding reward is; design actions to improve the highest-priority components. Vary topics and partners to support learning and social diversity.

Return ONLY the bullets.

Plan Prompt

System: You are {name}, a user participating in a multi-round social media game. In each round you will execute exactly {n_actions} actions and your choices accumulate

across rounds, affecting future relationships, topics seen and rewards. Recall past interactions and relationships from your memory when planning your actions.

Your best was round {best_round} with score {best_reward} when you did {best_summary}.
Your only goal is to grow what you value most-{active_comps}-in these proportions: {active_weights}.
[You got this coaching tip: {tip}]

Here are the reward definitions and how they translate into behaviours: {selected_reward_docs}
Here's how they combine:
Your total score is a mix of these parts (SOC, INF, PRE, COORD, EMO), weighted by λ's that sum to 1. Each λ decides a component's importance. Current weights: {active_weights}

Reward-aware rules (apply when the corresponding λ is large):
· COORD: For POST/COM, set "mention_flag": true when you plan to include a literal @mention (e.g., @post_author or a user from Users).
· SOC: Balance COM/DM between initiating and replying; if someone messaged you last round, reply.
· INF: Prefer topics not seen recently; otherwise keep a diverse mix from Topics.
· PRE: Include at least one POST (keep it concise and end with a clear question).
· EMO: Set "tone": "supportive" for COM/DM to elicit supportive replies.

PLANNING CONTRACT:
Plan EXACTLY {n_actions} actions.
When coach tips are provided, you MUST follow them in your plan. Only adapt if a tip is impossible to execute.
If a tip is impossible, replace only that action with the closest feasible alternative that preserves the tip's intent.
Respond with a raw JSON array only (no code fences, no comments, no extra keys, no trailing commas).

Each action must be an object with EXACTLY these keys:
· "type": one of "POST","COM","DM","NOT"
· "recipient": user ID for DM, otherwise null
· "topic": string for POST/COM/DM, null for NOT
· "target_id": COM → a post ID from the list; DM → a DM reply ID (reply mode) or null (cold DM); null for POST and NOT
· "mention_flag": boolean; true only if you will include an @mention (POST/COM only; must be false for DM and NOT)
· "tone": one of "supportive","neutral","critical"

CONSTRAINTS:
· POST: recipient=null, target_id=null
· COM: recipient=null, target_id MUST be a valid post ID from the list
· DM: two modes:
    - Reply DM: recipient MUST be the author of a listed DM reply AND target_id MUST be that DM id
    - Cold DM: recipient MUST be a valid user ID from the list AND target_id MUST be null
· NOT: recipient=topic=target_id=null

TOPIC GUIDANCE:
· For COM: use the target post's topic if provided; otherwise pick a relevant engaging topic.
· For POST/DM: choose any topic likely to engage the audience or recipient; exploring new topics can increase INF reward.
· There is no fixed topic list; you may introduce new topics, but keep them $\leq$ 5 words.

User: Last actions (most recent round): {last_actions},
Last observed reward scores: {observed_rewards},
Users you can interact with: {users},
Known topics so far: {topics} (you may introduce new topics),
DM replies (users who DM'd you last round):

13

```
- author: {dm_author} (DM id: {dm_id}, topic: {dm_topic}, content: {dm_content})
- ...
Commentable posts (choose a target_id from this list ONLY):
- target_id:  {p_id} (POST author:  {p_author},  topic:  {p_topic},  content:
  {p_content})
- ...
Based on the above context, decide on {n_actions} actions that best improve your
priority rewards. For Reply DM, use the listed DM IDs; cold DM can be sent to any
user in the list. For COM, you must supply a valid post ID from the list above.
Remember: replies help SOC, mentions help COORD, and supportive tone helps EMO.
```

Finally, we present the *Vote* prompt, a key component of the social-media simulation that captures how each agent, given its personality parameters, evaluates other users' content. The *Vote* prompt allows a user to express either liking or disliking of exposed content.

Vote Prompt

```
System: You are {name}, a user in a multi-round social media game.  You are a
{persona_type} with {personality} personality - use these traits to guide your
voting.
Your choices to "like" (1), "dislike" (-1), or remain neutral (0) on each piece
of content affect how others perceive you (PRE) and the emotional tone of your
interactions (EMO), and therefore influence your future rewards. Use your persona
traits and your relationship history to decide which content to support, oppose,
or ignore. A neutral vote (0) means you have no strong opinion or the relationship
context is neutral.

Respond ONLY with a JSON array of objects. Each object must have:
· id: the content ID (integer)
· vote: 1 for "like", -1 for "dislike", or 0 for "no vote"
Example output:
[
"id": 42, "vote": 1,
"id": 73, "vote": -1,
"id": 99, "vote": 0
]
No extra keys, no free-form text, no explanation.

User: Here are the latest items to vote on (id, sender, content{, relationship}):

      · {id} {sender} {content} {relationship}

      · {id} {sender} {content} {relationship}

      · {id} {sender} {content} {relationship}

      · ...

Vote on each according to your persona and the context of your relationships.
Remember: liking or disliking influences your reputation and future interactions;
use 0 ("no vote") when you have no strong opinion or the relationship context is
neutral.
```

The *Tie-Update* prompt is used to reweigh social ties after each round. It presents the peers interacted with, along with the corresponding text.

Tie-Update Prompt

```
System: You are {name}, a user in a multi-round social media game. Your task is to
judge the strength of evidence for increasing your social connection to each peer
based SOLELY on the interactions in the latest round.

SCORING GUIDE: Assign an integer score from 0 to 5 to each peer based on these
criteria:
· 5: Exceptional – repeated warmth/help/coordination; clear constructive alignment.
```

```
· 4: Strong — mutual positivity or clear support/assistance.
· 3: Good — polite/positive tone with some constructive exchange.
· 2: Weak — minor positive cues; limited substance.
· 1: Very Weak - faint positivity; likely noise.
· 0: None - mixed/negative/insufficient; do NOT increase.

RULES:
· Base decisions on LAST-ROUND transcript only; do not infer beyond text.
· Be conservative; if unsure, choose 0.
· One rating per peer. Reason must be factual and ≤1 sentence

OUTPUT FORMAT: You MUST output STRICT JSON. Include a score and a concise reason
for every peer provided.
{{
"ratings": [
     {{
     "peer": "PeerName",
     "score": 0|1|2|3|4|5,
     "reason": "≤ 1 sentence."
     }}
]
}}

User: Here are your peers for this round: {peer_list}.
Transcript of the last round: {transcript_text}.
Rate each peer according to the rules above.
```

B   Extended Formulation of the Social Tie Mechanism

In Sec. 3.4, we outlined the main mechanism by which directed social ties evolve over time as agents interact across rounds. That formulation left implicit the precise logic used to determine whether a directed interaction from user v to user u should be considered active in a given round. Moreover, we did not formally define the logic underlying the evidence score. We now address both points by first detailing the interaction activation criteria.

Interaction Activation.   Each action $a \in \mathcal{B}_t$ is annotated with structured metadata, including the action type $\text{TYPE}(a) \in \{\text{DM}, \text{COM}, \text{POST}, \text{NOT}\}$, the sender $\text{SNDR}(a) \in \mathcal{V}$, the recipient $\text{RCPT}(a) \in \mathcal{V}$ (for direct messages), and the set of mentioned users $\text{MENTIONS}(a) \subseteq \mathcal{V}$. Based on this metadata, we define a series of logical conditions to detect directed interactions between users.

To determine whether user v has sent a direct message to user u, we define

$$\text{isDmTo}(a, u, v) \coloneqq \big(\text{TYPE}(a) = \text{DM}\big) \wedge \big(\text{SNDR}(a) = v\big) \wedge \big(\text{RCPT}(a) = u\big).$$

To identify whether user v mentioned user u in a public message—such as a post or a comment—we define:

$$\text{isMention}(a, u, v) \coloneqq \big(\text{TYPE}(a) \in \{\text{POST}, \text{COM}\}\big) \wedge \big(\text{SNDR}(a) = v\big) \wedge \big(u \in \text{MENTIONS}(a)\big).$$

These two conditions jointly define the ADDRESS interaction channel, which is triggered when user v explicitly addresses user u through either a private message or a public mention.

To define the ENGAGE channel—capturing interactions with content authored by user u—we rely on two additional metadata fields. The field $\text{TARG}(a) \in \mathcal{V}$ denotes the author of the content that action a targets. Based on this, we define whether user v comments on content created by u as

$$\text{isComTo}(a, u, v) \coloneqq \big(\text{TYPE}(a) = \text{COM}\big) \wedge \big(\text{SNDR}(a) = v\big) \wedge \big(\text{TARG}(a) = u\big).$$

To capture positive or negative evaluation of content, we include voting behavior. The field $\text{VOTE}(a) \in \{-1, 0, +1\}$ indicates a downvote, no vote, or an upvote, respectively. We define

$$\text{isVoteFor}(a, u, v) \coloneqq \big(\text{VOTE}(a) \neq 0\big) \wedge \big(\text{SNDR}(a) = v\big) \wedge \big(\text{TARG}(a) = u\big).$$

Together, these logical definitions specify whether a directed interaction from user v to user u is considered active in round t. Specifically, the binary activation variable $\zeta_t(v \to u)$ is set to 1 if either the ADDRESS or ENGAGE channel is triggered by any action in $\mathcal{B}_t$.

Evidence Score. The evidence score $e_t(v \to u)$ aggregates four distinct dyadic signals that capture different dimensions of social interaction between users v and u.

Novelty. This signal quantifies the introduction of new information from v to u. Let $\mathcal{T}_t(v \to u)$ denote the set of topics introduced by v to u during round t, and let $\mathcal{H}_{t-1}^\tau(u)$ represent u's historical topic exposure up to time $t - 1$. The novelty signal is defined as:

$$S_t^{\mathrm{Nov}}(v \to u) = \mathbb{1}\{\exists \tau \in \mathcal{T}_t(v \to u) : \tau \notin \mathcal{H}_{t-1}^\tau(u)\}$$

which indicates whether v introduced any topics novel to u's experience.

Approval. This signal captures explicit evaluative feedback from v on u's content. Let $L_t^+(v \to u)$ and $L_t^-(v \to u)$ denote the counts of likes and dislikes, respectively, that v assigned to u's content during round t. The approval signal is computed as:

$$S_t^{\mathrm{Appr}}(v \to u) = \begin{cases} 0, & L_t^+(v \to u) + L_t^-(v \to u) = 0, \\ \dfrac{L_t^+(v \to u) - L_t^-(v \to u)}{\max\{1,\ L_t^+(v \to u) + L_t^-(v \to u)\}}, & \text{otherwise.} \end{cases}$$

This measures the net approval normalized by total engagement, ranging from -1 to 1.

Reciprocity. This signal quantifies the balance in positive engagement between users over time. Using an exponential moving average to smooth temporal fluctuations:

$$\hat{L}_t^+(v \to u) = \beta^{\mathrm{rec}} \hat{L}_{t-1}^+(v \to u) + (1 - \beta^{\mathrm{rec}}) L_t^+(v \to u)$$

where $\beta^{\mathrm{rec}} \in (0, 1)$ controls the memory persistence. The reciprocity signal then measures the symmetry in smoothed like exchanges:

$$S_t^{\mathrm{rec}}(v \to u) = 1 - \left| \frac{\hat{L}_t^+(v \to u) - \hat{L}_t^+(u \to v)}{\hat{L}_t^+(v \to u) + \hat{L}_t^+(u \to v) + \epsilon} \right|$$

where $\epsilon > 0$ ensures numerical stability. This formulation yields values near 1 for balanced relationships and near 0 for highly asymmetric engagement.

Affective Tone. This signal captures the emotional valence of private communication when explicit feedback is absent. Let $s(a) \in [-1, 1]$ represent the sentiment score of action a. The affective tone signal is defined as:

$$S_t^{\mathrm{Aff}}(v \to u) = \begin{cases} \mathrm{avg}\{ s(a) : a \in \mathcal{B}_t,\ \mathrm{SNDR}(a) = v,\ \mathrm{RCPT}(a) = u,\ \mathrm{TYPE}(a) = \mathrm{DM}\}, & \text{if } L_t^+ + L_t^- = 0 \wedge \exists\, \mathrm{DM} \\ 0, & \text{otherwise.} \end{cases}$$

This ensures that emotional tone is only considered when no explicit evaluations (likes/dislikes) are present, capturing the unique contribution of affective communication.

C   Social Network Characteristics

C.1   Evaluation Metrics

In accordance with the network science literature (see [5]), we define the following standard notation relevant to the analysis of social networks.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ be a (possibly weighted) *directed* network with $|\mathcal{V}|$ nodes and adjacency matrix $A \in \mathbb{R}_{\geq 0}^{|\mathcal{V}| \times |\mathcal{V}|}$, with

$$A_{uv} = \begin{cases} > 0, & \text{if } (u \to v) \in \mathcal{E}, \\ 0, & \text{otherwise,} \end{cases}$$

and let the out-degree of node u be defined by $k_u^{\mathrm{out}} = \sum_v A_{uv}$, and the in-degree of node u be given by $k_u^{\mathrm{in}} = \sum_v A_{vu}$. When $\mathcal{G}$ is *undirected*, the matrix A is symmetric, with $A_{uv} = A_{vu}$, and we write simply $k_u = \sum_v A_{uv}$.

Degree distribution. The out-degree sequence $\{k_u^{\mathrm{out}} : u \in \mathcal{V}\}$ induces the empirical probability mass function

$$P(k) = \frac{\left|\{u : k_u^{\mathrm{out}} = k\}\right|}{|\mathcal{V}|}.$$

Alternatively, one may work with the normalized degree $\tilde{k}_u = k_u^{\mathrm{out}}/(|\mathcal{V}| - 1) \in [0, 1]$, and analyze its distribution via the scaling relation $P(\tilde{k}) \propto \tilde{k}^{-\gamma}$, with the exponent γ estimated through statistical inference on the empirical distribution. By the same logic, the in-degree distribution can be determined.

Density. Assume the edge weights satisfy $w : \mathcal{E} \to [0,1]$. The density of the directed network $\mathcal{G}$ is defined as

$$\rho(\mathcal{G}) = \frac{\sum_{u \neq v} A_{uv}}{|\mathcal{V}|\,(|\mathcal{V}| - 1)},$$

which lies in $[0,1]$ and measures the average weight across all possible directed edges. In the unweighted case, the numerator simplifies to $|\mathcal{E}|$, the number of directed edges. In the undirected case, the analogous quantity is

$$\rho(\mathcal{G}_{\mathrm{ud}}) = \frac{2 \sum_{u < v} A_{uv}}{|\mathcal{V}|\,(|\mathcal{V}| - 1)},$$

where the factor of 2 ensures normalization, and $\rho(\mathcal{G}_{\mathrm{ud}})$ similarly lies in $[0,1]$.

Clustering Coefficient. For each vertex $u \in \mathcal{V}$ in the directed network $\mathcal{G}$, assume $A_{uu} = 0$ (no self-loops). Let the out-degree be $k_u^{\mathrm{out}} = \sum_{v \in \mathcal{V}} A_{uv}$. The local directed clustering coefficient, which quantifies the likelihood that the out-neighbors of u are interconnected in a directed sense, is defined by

$$\kappa_u^{\mathcal{G}} = \frac{\sum_{\substack{v,v' \in \mathcal{V} \\ v \neq v'}} A_{uv}\, A_{vv'}\, A_{v'u}}{k_u^{\mathrm{out}}\,(k_u^{\mathrm{out}} - 1)},$$

and the global clustering coefficient is $\kappa(\mathcal{G}) = (1/|\mathcal{V}|) \sum_{u \in \mathcal{V}} \kappa_u^{\mathcal{G}}$. In the undirected network $\mathcal{G}_{\mathrm{ud}}$, where $A_{uv} = A_{vu}$, let $k_u = \sum_{v \in \mathcal{V}} A_{uv}$ and define the neighborhood $\mathcal{N}(u) = \{v \in \mathcal{V} : A_{uv} > 0\}$. Let $\mathcal{E}_u = \big|\{\{v,v'\} \subset \mathcal{N}(u) : A_{vv'} > 0\}\big|$ denote the number of edges between neighbors of u. The local undirected clustering coefficient is then

$$\kappa_u^{\mathcal{G}_{\mathrm{ud}}} = \frac{2\,\mathcal{E}_u}{k_u\,(k_u - 1)},$$

and the global clustering coefficient is $\kappa(\mathcal{G}_{\mathrm{ud}}) = \frac{1}{|\mathcal{V}|} \sum_{u \in \mathcal{V}} \kappa_u^{\mathcal{G}_{\mathrm{ud}}}$. The clustering coefficient quantifies the tendency of the network to form triangles, providing insight into local connectivity patterns that complement degree-based summaries.

Largest Weakly Connected Component. Given a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $\mathcal{G}_{\mathrm{ud}} = (\mathcal{V}, \mathcal{E}_{\mathrm{ud}})$ be its underlying undirected graph $\mathcal{E}_{\mathrm{ud}} = \big\{\{u,v\} : (u \to v) \in \mathcal{E} \text{ or } (v \to u) \in \mathcal{E}\big\}$. A weakly connected component is a maximal subset $\mathcal{X} \subseteq \mathcal{V}$ in which every $u, v \in \mathcal{X}$ is joined by a path in $\mathcal{G}_{\mathrm{ud}}$. Denote by $\mathcal{X}_{\max}$ the largest such component; its relative size is

$$\mathrm{LCC}(\mathcal{G}) = \frac{|\mathcal{X}_{\max}|}{|\mathcal{V}|},$$

computed in $O(|\mathcal{V}| + |\mathcal{E}|)$ time using any standard linear-time algorithm for connected components.

Average Shortest Path Length. Restrict to the largest weakly connected component $\mathcal{X}_{\max}$. Define $d(v, u)$ as the length of a shortest directed path from v to u, with the convention $d(v, u) = \infty$ if no such path exists. By averaging only over reachable pairs, we obtain the typical directed-path length within the LCC,

$$\mathcal{L}(\mathcal{G}) = \frac{\sum_{v \neq u} d(v,u)\,\mathbb{1}\{d(v,u) < \infty\}}{\sum_{v \neq u} \mathbb{1}\{d(v,u) < \infty\}},$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

Community Modularity. Given a partition $g^{(Q)} : \mathcal{V} \to \{1, \ldots, C^{(Q)}\}$ of the nodes into $C^{(Q)}$ communities, and denoting the total edge-weight by $E = \sum_{u,v} A_{uv}$, the modularity of $\mathcal{G}$ with respect to $g^{(Q)}$ is defined as

$$Q(\mathcal{G}, g^{(Q)}) = \frac{1}{E} \sum_{u,v=1}^{|\mathcal{V}|} \left(A_{uv} - \frac{k_u^{\mathrm{out}}\, k_v^{\mathrm{in}}}{E} \right) \delta\big(g^{(Q)}(v), g^{(Q)}(u)\big),$$

where δ denotes the Kronecker delta function. For clarity, in the unweighted case, $E = 2|\mathcal{E}|$ for undirected graphs and $E = |\mathcal{E}|$ for directed graphs.

Homophily. Let $g^{(\phi)} : \mathcal{V} \to \{1, \ldots, C^{(\phi)}\}$ assign each node to one of $C^{(\phi)}$ groups, and denote by $|\mathcal{V}_r| = |\{i : g^{(\phi)}(i) = r\}|$ the size of group r, such that $\sum_{r=1}^{C^{(\phi)}} |\mathcal{V}_r| = |\mathcal{V}|$. Denoting the total edge-weight by $E = \sum_{u,v} A_{uv}$, we define the observed cross-group weight as

$$\widehat{W}(\mathcal{G}, g^{(\phi)}) = \sum_{u,v=1}^{|\mathcal{V}|} A_{uv}\, \mathbb{1}\{g^{(\phi)}(u) \neq g^{(\phi)}(v)\},$$

and the expected cross-group weight under random mixing as

$$\mathbb{E}[W(\mathcal{G}, g^{(\phi)})] = E\Big(1 - \sum_{r=1}^{C^{(\phi)}} \big(\tfrac{|\mathcal{V}_r|}{|\mathcal{V}|}\big)^2\Big).$$

The homophily function ϕ is then defined by

$$\phi(\mathcal{G}, g^{(\phi)}) = \frac{\widehat{W}(\mathcal{G}, g^{(\phi)})}{\mathbb{E}[W(\mathcal{G}, g^{(\phi)})]},$$

where $\phi(\mathcal{G}, g^{(\phi)}) < 1$ indicates homophily (fewer cross-group ties than expected under random mixing), $\phi(\mathcal{G}, g^{(\phi)}) = 1$ indicates random mixing, and $\phi(\mathcal{G}, g^{(\phi)}) > 1$ indicates heterophily. Beyond topology-based homophily, homophily can also be operationalized behaviorally by first inferring user policies and then comparing them across users; see [42]. Such comparisons allow for assessing a reward-related form of homophily.

C.2 Degree Distributions

In Sec. 4, we presented a set of network characteristics for graphs obtained from the multi-agent LLM simulation. For the coach evaluation, we primarily relied on reward maximization and average network statistics as evaluation metrics for the tie mechanism. Here, we further examine the degree distributions of the final undirected graphs $\mathcal{G}_T$ after applying the conversion threshold θ. Figure 5 shows the degree distributions under the heuristic tie mechanism without the coach in the planning process, whereas Fig. 6 presents the same mechanism with the coach. Likewise, Fig.7 reports the distributions under the LLM text-based tie mechanism without the coach, and Fig. 8 with the coach in the tie reweighting process. The results indicate that the presence of the coach increases the median degree $\tilde{k}$ of the resulting networks, suggesting that coaching induces agents to maintain more active connections on average. The text-based tie formation mechanism yields consistently higher median degrees than the heuristic mechanism across all threshold values, indicating denser and more robust network structures.
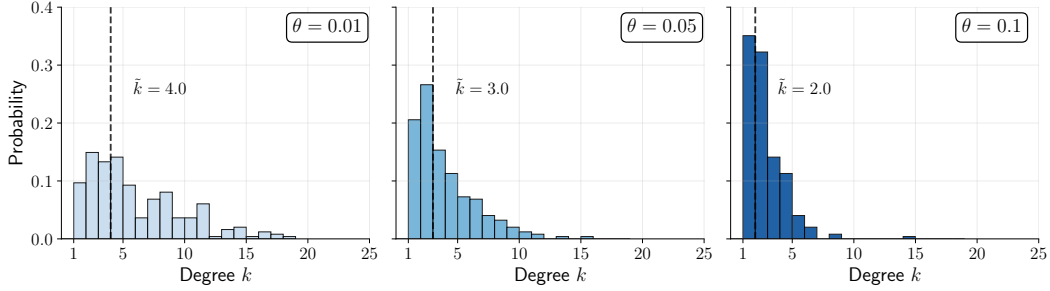

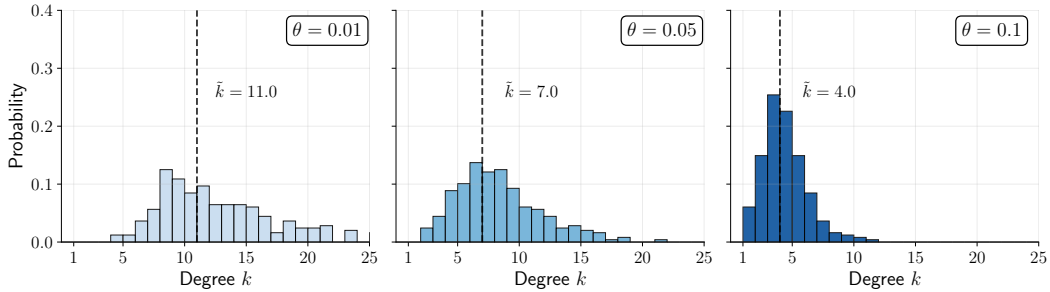Figure 5: Degree distributions for heuristic tie formation without coach.

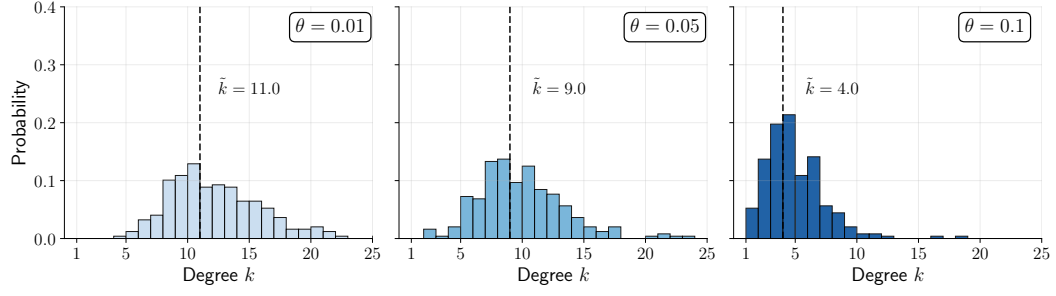Figure 6: Degree distributions for heuristic tie formation with coach.

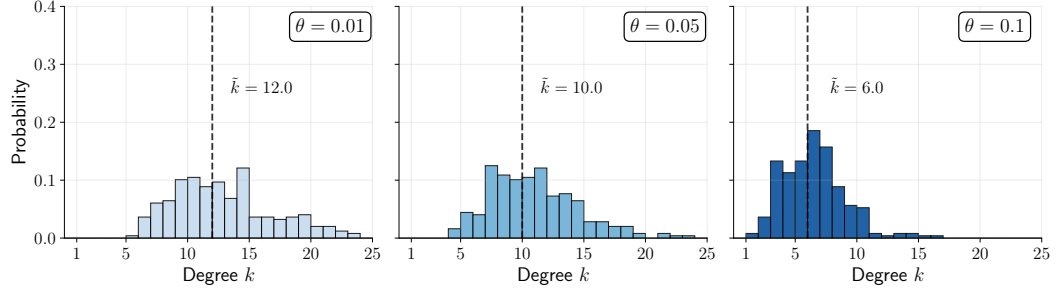Figure 7: Degree distributions for LLM text-based tie formation without coach.

Figure 8: Degree distributions for LLM text-based tie formation with coach.