Modeling Network Formation with LLM Agents: The Role of Demographics and Personality

Christos Gkartzios
Department of Computer Science & Engineering
University of Ioannina
Ioannina, Greece
chgartzios@cs.uoi.gr

Evaggelia Pitoura
Department of Computer Science & Engineering
University of Ioannina
Ioannina, Greece
pitoura@uoi.gr

Panayiotis Tsaparas
Department of Computer Science & Engineering
University of Ioannina
Ioannina, Greece
tsap@uoi.gr

***Abstract*—Networks are central to many applications, yet the mechanisms underlying their formation are not fully understood. In this paper, we investigate network formation using LLM agents, focusing on whether demographics and personality influence connection choices. We introduce a simple network formation process in which, at each timestep, a candidate pool is constructed for each agent from its two-hop neighbors and a set of randomly sampled nodes, independently of agent attributes. By systematically controlling which attributes of the candidates are revealed to the agent, we study the effect of demographics and personality traits on network formation.**

Our findings indicate strong homophily across multiple dimensions, particularly demographics. We further observe that popularity (i.e., number of connections) is shaped by personality traits, such as extraversion. In addition, we analyze the effects of reciprocity and temporal dynamics on these patterns. Our results have dual significance: they shed light on the mechanisms driving social network formation, while also revealing potential biases embedded in LLM models.

***Index Terms*—Large Language Models (LLMs); Network formation; Social networks; Demographics and Personality traits; Homophily**

I. Introduction

Networks capture relationships across social platforms, professional settings, code collaboration, and citation systems. Structure matters for who sees what. Acquaintances open new routes, similarity pulls interactions inward, and platform mechanisms such as ranking, recommendation feeds, and trending modules direct attention [1]–[3]. These dynamics are evident on major platforms. On Facebook the social graph shows high clustering and short paths, and over time user interactions concentrate on a small core of contacts [4], [5]. On Twitter influence is heavy tailed, so a small share of accounts drive most retweets and mentions [6]. On GitHub transparency in activity and code history supports coordination and reputation driven collaboration [7].

Agent-based modeling (ABM) links micro decisions to macro patterns and lets us study these dynamics in a controlled way. Heterogeneous agents follow simple local rules and generate aggregate patterns such as consensus, clustering, or polarization [8]–[11]. Recent systems pair this ABM design with language model agents that perceive context, store short memories, and act via text, yielding multi step social behavior in open environments [12]–[14]. In parallel, LLM simulations on minimal platforms reproduce attention inequality and echo chambers, and probe how feed rules shape outcomes [15], while other work fixes the network and studies belief dynamics [16]. We take a complementary route focused on *link formation*: we keep a small, explicit link-formation rule and let LLM agents make selection and acceptance decisions.

Specifically, we study graph creation with an explicit link formation rule driven by LLM agents. Each source draws a candidate pool from local structure, selects targets as discrete choice [17], and targets may accept or reject. We vary visibility of demographics and personality traits and report standardized graph metrics with matched generative baselines [18]–[21]. This contrasts with prior work that either ties growth to content engagement [15], fixes the interaction network to study belief dynamics [16], or supplies engineered network features for link choice [22]. This setup lets us examine effects of attribute visibility and bilateral acceptance on connectivity, community structure, and homophily over time.

We start from an initially empty network so that structure is an outcome rather than an input, which removes inherited topology and makes the role of information and acceptance directly observable.

We hold fixed the language model, prompt texts, and per-step proposal budget, and vary only two parameters: whether node attributes are visible to agents and whether targets can accept or reject proposals. This isolates the role of attribute visibility and bilateral acceptance in shaping the resulting network. We organize the evaluation around four questions. First, homophily under visibility: how revealing demographics or personality traits changes same-attribute links and community structure. Second, connectivity: how demographics and personality traits shape selection behavior and degree. Third, reciprocity: how requiring the target's acceptance changes which proposals become edges and the resulting structure. Fourth, time: how homophily and connectivity evolve across timesteps.

We find that making demographics visible shifts selection toward same-group choices and strengthens community structure. Making personality traits visible increases trait-based homophily and introduces cross-cutting signals that attenu-

ate purely demographic effects. Requiring target acceptance reduces realized edges, increases separation, and modestly amplifies homophily. Degree associates positively with extraversion and weakly with agreeableness, and negatively with neuroticism. Relative to random and preferential-attachment baselines matched on size and density, the resulting networks exhibit clearer meso-scale structure driven by agent decisions.

Contributions.

- **A transparent algorithm for LLM mediated, agent driven graph creation.** At each step, agents build a small candidate pool from friends of friends (FoF) plus k uniformly sampled non-neighbors, select up to k targets, record rejections in a persistent memory $M_t(a)$, and add timestamped edges. The procedure is simple enough to inspect end to end.
- **Two parameter design isolating visibility and acceptance.** We vary only two parameters, whether demographics and personality traits are visible and whether targets must accept, while holding the model, prompts, and the per-step number of connection attempts fixed. This lets us attribute structural changes to these parameters.
- **Graph-centric evaluation with baseline comparisons.** We focus on three areas: community structure (Louvain, modularity), connectivity (degree) and homophily, and reciprocity (effects of acceptance). We compare against Erdős–Rényi and Barabási–Albert graphs matched on size and density. Temporal evolution is included as snapshots at t=1, 2, 3.

The remainder of the paper is structured as follows. Sec. II presents the agent-based link-formation rule. Sec. III details the experimental evaluation, including research questions (Sec. III-A), setup/metrics (Sec. III-B) and evaluation results (Sec. III-C). Sec. IV reviews related work. Sec. V discusses limitations and next steps. Sec. VI concludes. Prompt texts appear in the Appendix.

II. Network Creation Algorithm

Networks play a central role in many domains, yet the principles that govern their formation remain insufficiently understood. In this paper, we consider an initially empty network of *LLM agents*, and examine the types of networks that emerge when connections are formed by the agents over time.

Specifically, agents are assigned attributes that remain fixed throughout the process. At each timestep, an agent selects to connect to another agent from a *candidate pool* of agents. For the selection, we control the information provided to the source agent regarding the attributes of the agents in the pool. For reciprocity, we also model cases where the target agent may reject the connection request. Our overall goal is to understand how attributes shape edge formation and, in turn, the evolution of connectivity, community structure, and homophily in the network.

Algorithm 1 Agent Creation

```
Input: number of agents n
Output: set A with n new agents
for i = 1 to n do
    Create agent a
    Assign demographic attributes to a
    Assign Big Five traits to a
    Initialize rejection memory M_0(a) ← ∅
    Append a to A
```

1) Agent creation: Let $G_t = (V, E_t)$ denote the cumulative undirected graph after timestep t, where V is the set of agents and E_t the set of edges where each edge is a timestamped tuple (i, j, τ), $i, j \in V$, and $\tau \leq t$. Initially, E_t is empty.

The process of agent creation is described in Algorithm 1. Each agent $a \in V$ is assigned attributes specifying its demographics and personality.

Each agent is first assigned demographic attributes drawn from predefined distributions. The three attributes considered are sex, race, and age group, as summarized in Table I, adapted from U.S. Census Bureau American Community Survey (ACS) categories.[1] Sex and race are drawn independently from fixed categorical distributions that approximate population frequencies. Age group is assigned by mapping a normal distribution over adult ages onto the buckets shown in Table I.

In addition to demographic attributes, each agent is assigned *Big Five trait scores* (OCEAN) [23], [24]: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These traits provide a continuous representation of individual differences in behavior and social preferences. In our simulation, values for each trait are sampled from a truncated normal distribution in $[0, 1]$, independent of demographics, to ensure variation while remaining within plausible ranges.

TABLE I: Demographic attribute values used in simulation

Attribute	Values
Sex	Male, Female
Race	White, Black, Asian, Hispanic
Age Group	18–24, 25–34, 35–44, 45–54, 55–64, 65+

Demographic and psychological attributes together serve as inputs for connection decisions. Demographics ensure that population diversity is explicitly represented in the simulation, while the Big Five traits introduce heterogeneous decision-making patterns, influencing both the number of connections agents attempt and the likelihood of accepting or rejecting connection requests.

Finally, each agent a maintains a persistent rejection memory $M_t(a) \subseteq V \setminus \{a\}$ that stores the identities of agents that declined its connection requests in the past.

2) Network creation: The process of connection creation is specified in Algorithm 2.

[1] See https://www.census.gov.

Algorithm 2 Network Creation

Input: number of agents n, timesteps T, demographic/trait distributions, model
Output: timestamped undirected graph $G_t = (V, E_t)$ with node attributes
Initialize agent list $\mathcal{A} \leftarrow []$
AgentCreation(n) ▷ create n agents with attributes
Initialize graph G_0 with $V \leftarrow \mathcal{A}$, $E_0 \leftarrow \emptyset$
for $t = 1$ to T **do**
  for each agent $a \in \mathcal{A}$ **do**
    Fixed number of new connections k for a
    Initialize $C_t(a)$ with second-hop nodes at t
    Augment $C_t(a)$ with k uniformly sampled non-neighbors of a
    Exclude current neighbors $N_t(a)$
    Exclude prior rejectors $M_t(a)$ from $C_t(a)$
    $S_t(a) \leftarrow$ model selects up to k targets from $C_t(a)$
    for each $b \in S_t(a)$ with $(a, b) \notin E_t$ **do**
      if an acceptance step is required **then**
        query b for ACCEPT/REJECT
        if ACCEPT **then**
          add (a, b, t) to E_t
        else
          add b to $M_t(a)$
      else
        add (a, b, t) to E_t
Save node attributes and timestamped edge list

At $t = 0$, we create n agents and assign to them fixed attributes that persist for the entire simulation.

At each timestep t, each agent a initiates k connection attempts from a candidate pool of agents. The candidate pool for agent a consists of all second-hop (friend-of-friend) nodes, plus k uniformly sampled nodes, excluding the self node a, current neighbors $N_t(a)$, and prior rejectors $M_t(a)$.

Specifically, for agent a, let

$$N_t(a) = \{\, u \in V : \ (a, u) \in E_t \,\}$$

be its neighbor set at time t.

We define the friend-of-friend set as

$$\mathsf{FoF}_t(a) = \big\{\, u \in V \setminus \big(N_t(a) \cup \{a\}\big) : \\ \exists\, v \in N_t(a) \text{ such that } (u, v) \in E_t \,\big\}.$$

Agent a maintains a persistent rejection memory $M_t(a) \subseteq V$ with the identities of agents that rejected a in the past. If a proposes to b at step t and b rejects, we update $M_t(a) \leftarrow M_t(a) \cup \{b\}$. The memory records identities only and does not include time, context, or reasons. At every step the candidate pool excludes prior rejectors.

Let $U_t(a)$ be a size-k sample drawn uniformly without replacement from $V \setminus \big(N_t(a) \cup \{a\} \cup M_t(a)\big)$, then the candidate pool can be written as

$$C_t(a) = \big(\mathsf{FoF}_t(a) \ \cup \ U_t(a)\big) \setminus M_t(a).$$

The model then selects up to k targets $S_t(a) \subseteq C_t(a)$. We use the same k for sampling and selection.

3) Prompt design: We now provide details about the two templates we use: a *selection* prompt and an *acceptance* prompt. Both prompts are parameterized by the current timestep t, the focal agent a, and the candidate pool $C_t(a)$.

- **Selection prompt.** The prompt includes: (1) a header with agent a's own demographics and Big Five traits (2) a list of candidates $u \in C_t(a)$ that includes an anonymized ID (for example, `U17`), whether u is a friend of friend (FoF) or a random draw and any attributes, and (3) the instruction that asks for up to k different targets from $C_t(a)$ in a fixed output format.
- **Acceptance prompt.** The prompt includes: (1) a header with target b's own demographics and Big Five traits, (2) a initiator summary that shows a's anonymized ID (e.g., `U17`), whether a is a FoF or a random draw relative to b and any attributes, and (3) the instruction that requests a binary decision (ACCEPT or REJECT) with a one-sentence rationale. If the decision is REJECT, b is added to $M_t(a)$ so that a does not propose to b again.

All runs use a fixed prompt template for reproducibility.

III. Evaluation

We examine how demographic attributes and Big Five traits shape the selection of connections. We evaluate an agent driven process in which each agent selects a target from a candidate pool, and, when acceptance is on, the selected target decides whether to accept the proposal. To assess the role of agent characteristics in link selection, we vary what is visible to the agents by revealing or hiding demographics (age, sex, race) and Big Five traits. To evaluate the impact of reciprocity, we toggle acceptance so that a proposed edge is added only if the target agrees. We run the selection process in multiple timesteps to see how connections evolve. We organize the evaluation around four questions on visibility, connectivity, reciprocity, and temporal evolution.

The exact agent persona, candidate–selection, and acceptance prompts are provided in the Appendix.

All experiments ran on a dual socket AMD EPYC 9335 system with 64 cores and 128 threads, ∼1.0 TiB RAM, and two NVIDIA H200 NVL GPUs with 143,771 MiB memory.

A. Research Questions

RQ1 **Homophily.** How does visibility influence same attribute connections and community structure across demographics and Big Five traits?

RQ2 **Connectivity.** How do demographics or Big Five traits affect selection behavior and node degrees in the resulting network?

RQ3 **Reciprocity.** What is the impact of acceptance on realized connections and how does it shape the emergent network structure?

RQ4 **Temporal evolution.** How do homophily and connectivity evolve in time?

TABLE II: Parameters used in evaluation

Symbol	Setting	Meaning
A0	Acceptance off	All proposed edges are added
A1	Acceptance on	Target accepts or rejects
D0	Demographics off	No demographic attributes
D1	Demographics on	Demographics are visible during the selection process
B0	Big Five off	No Big Five attributes
B1	Big Five on	Big Five traits are visible during the selection process

B. Setup and Metrics

We evaluate how attributes shape target selection and connection formation. The process runs for three timesteps. At each timestep, every agent attempts $k = 3$ new connections from its candidate pool, and rejection memory persists across timesteps. All targeting and acceptance decisions use Qwen3 [25] with 30B parameters. Reported values are averages over five independent runs, and unless otherwise noted we report the final snapshot at $t = 3$ and compute all metrics on the largest connected component (LCC). We vary the visibility of demographics (**D0** and **D1**) and Big Five traits (**B0** and **B1**), and whether acceptance is required (**A0** and **A1**).

We study eight settings by toggling three parameters for acceptance **A**, demographics **D**, and Big Five attributes **B** (Table II). All agent decisions use the same model. Agents always know their own demographic and personality attributes.

All settings run on the same discrete timeline. For each agent a at step t, the candidate pool $C_t(a)$ is the union of all second hop nodes and k uniformly sampled nodes that are not current neighbors. The self node and prior rejectors $M_t(a)$ are removed from the pool. The rejection memory $M_t(a)$ persists across timesteps. In **A0** all proposed edges are added. In **A1** the target decides whether to accept or reject, and rejections are added to $M_t(a)$. When **D1** and/or **B1** are on, the corresponding attributes are available to agents and may influence both targeting and acceptance; under **D0/B0** they are withheld from decision prompts. Note that in all cases, an agent is aware of its own demographic and personality characteristics.

Homophily (RQ1) and **Connectivity (RQ2).** For homophily and connectivity we vary what agents can see. We compare **D0** to **D1** and **B0** to **B1** while keeping the mechanism fixed. We report same attribute connection rates, community count, modularity, degree patterns, and the size of the largest connected component to understand how attribute visibility affects target selection and community structure.

Reciprocity (RQ3). For reciprocity we compare **A0** and **A1** while holding **B** and **D** fixed. We seek to understand the role of acceptance in network formation and its impact on structure. We measure changes in edge count, average degree, diameter, community count, modularity, and homophily to isolate the effect of the target's decision on the network.

Temporal evolution (RQ4). Our goal is to understand how networks grow and how connectivity and community structure develop over time. We report metrics at $t = 1, 2, 3$ and examine the evolution of homophily, edge count, average degree, community count, and modularity between the first and final snapshots.

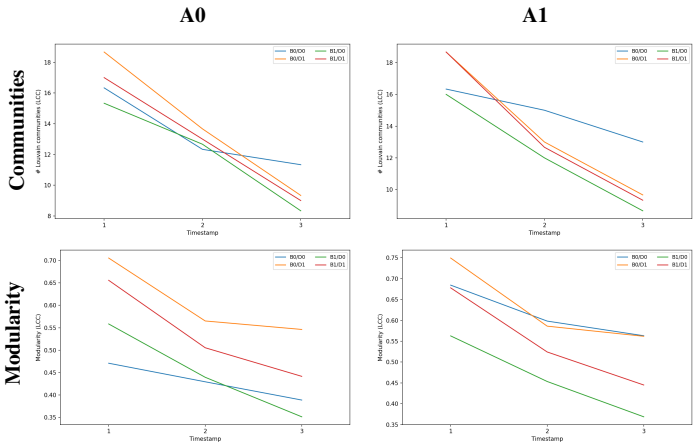

Fig. 1: Louvain Community structure over time on the Largest Connected Component (LCC).

The metrics used in the study are reported next. By default we use the largest connected component and report $t = 1, 2, 3$, treating $t = 3$ as the final snapshot. Reported values are averages over five independent runs.

Metrics.

- **Modularity** Q [26], [27]: Modularity of the communities on the largest connected component.
- **Community count** K: number of Louvain communities on the largest connected component.
- **Edge count** m: total number of edges.
- **Average degree** $\bar{d}$: mean node degree.
- **Density** [28]: fraction of actual edges among possible edges on the largest connected component.
- **Diameter** [28]: longest shortest path length on the largest connected component.
- **Largest component size** n: number of nodes in the largest connected component.
- **Demographic homophily** [2], [29]: percent of edges whose endpoints share race, sex, or age group.
- **Trait homophily** [2]: percent of edges whose endpoints fall in the same trait buckets. Each trait is partitioned into four buckets on $[0, 1]$.
- **Degree to trait correlation** ρ [30]: Spearman rank correlation between degree d and a trait score x. Rank based and scale free. $\rho \approx 1$ means higher trait scores tend to have higher degree, $\rho \approx -1$ the reverse, $\rho \approx 0$ no monotone relation.
- **Degree to demographic correlation** η [31]: correlation ratio of degree across groups g. Interpreted as the share of degree variance explained by group membership. $\eta \approx 1$ means degree differs strongly by group with most variance between groups. $\eta \approx 0$ means groups have similar degree and little variance is explained by group

membership.

We include baselines to help us distinguish effects driven by agent choices from patterns produced by simple generative processes.

Baselines. Two generative baselines that provide contrasting null models are chosen:

- **Erdős and Rényi (ER)** [32]. Edges appear independently with equal probability for every pair of nodes and capture random mixing.
- **Barabási and Albert (BA)** [33]. New nodes attach preferentially to higher degree nodes and produce skewed degree distributions with hubs.

Both baselines use the same node count n as the agent runs and the mean final edge count m across configurations so that densities match. All structural metrics are computed in the same way for baselines and evaluated settings and are reported in Table IV.

C. Evaluation Results

1) Homophily: Demographics visibility (**D1**) shifts selection toward same demographic attribute targets and produces large homophily gains by the final snapshot. When demographics are visible (**B0**, **D1**, **A0**) race goes from 26.99% to 51.46% and age from 19.65% to 58.95% compared to where they are not visible (**B0**, **D0**, **A0**) *(Table III, Fig. 2)*. These shifts are accompanied by higher modularity from 0.389 to 0.546 *(Table IV, Fig. 1)* at comparable n and lower density from 0.033 to 0.029 *(Table IV)*.

When *Big Five* traits are visible and demographics are hidden (**B1** with **D0**) under **A0**, selection shifts toward same trait pairs across all traits at the final snapshot. The largest changes are in Openness and Extraversion. Openness rises from 41.00% to 52.42% which is an increase of 11.42 percentage points (Table III, Fig. 3). Extraversion rises from 39.10% to 45.48% which is an increase of 6.38 points (Table III, Fig. 3). At the other end, Agreeableness rises modestly from 39.53% to 40.87% which is an increase of 1.34 points (Table III, Fig. 3). In descending order of absolute change, the traits are Openness, Extraversion, Neuroticism, Conscientiousness, Agreeableness. Relative to traits hidden at the same settings (**B0** with **D0** under **A0**), modularity shows a small decrease, Q from 0.389 to 0.351 (Table IV, Fig. 1). This indicates the existence of more cross community links, which distribute edges more evenly and lower Q.

When demographics and Big Five traits are both visible (**B1** with **D1**) under **A0**, each attribute family has a distinct effect. Same trait matching increases, for example relative to **B0** with **D1** Openness goes from 41.31% to 46.49% and Neuroticism from 43.11% to 49.89% (Table III). At the same time race and age homophily decrease and sex homophily increases, for instance race goes from 51.46% to 46.61% and sex from 45.45% to 51.95% (Table III, Fig. 2). Community cohesion is lower than with demographics alone, with Q from 0.546 under **B0** with **D1** to 0.442 under **B1** with **D1** (Table IV, Fig. 1), consistent with more cross community links. In short Big Five

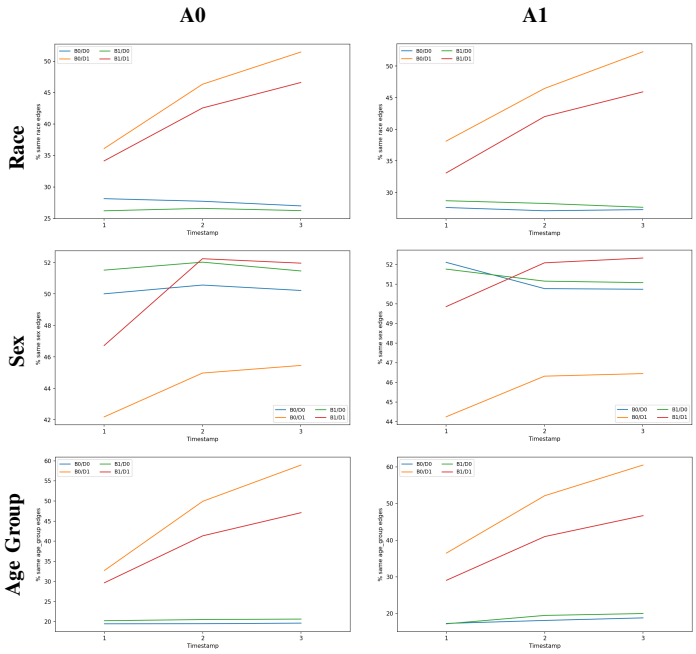

Fig. 2: Homophily for **demographic attributes**.

visibility introduces cross cutting structure that weakens race and age matching, strengthens same sex pairing, and reduces community cohesion compared to **D1** alone.

2) Connections: With traits visible and demographics hidden (**B1** with **D0**) under **A0**, degree increases with Extraversion and Agreeableness and decreases with Neuroticism (Fig. 5). This pattern appears at the first and final snapshots (Fig. 5). Extraversion shows the strongest positive association and Neuroticism the clearest negative one. At the graph level, average degree is similar across visibility settings, for example 16.565 under **B0** with **D0** and 15.656 under **B1** with **D0** at the final snapshot (Table IV).

For demographics, hiding attributes yields weak degree associations, and they become stronger when demographics are visible. Sex shows the most stable degree correlation while race and age display higher homophily but smaller or mixed degree correlations at the final snapshot (Fig. 4).

3) Reciprocity: Acceptance (**A1**) reduces the number of formed edges and increases separation. For example, under **B0** with **D0** edges go from 4141 to 3411, diameter from 4 to 5.33, modularity from 0.389 to 0.563, the number of communities from 11.33 to 13.00, and average degree from 16.565 to 13.725 (Tables IV, III, Fig. 1). When demographics are visible, **A1** increases demographic homophily slightly and consistently. For example, under **B0** with **D1** race goes from 51.46% to 52.25% and age from 58.95% to 60.53% (Table III). Trait homophily changes are small. With traits visible and demographics hidden under **B1** with **D0** Openness increases by 1.54 percentage points and Extraversion by 1.23 percentage points at the final snapshot (Table III). Under **A1**, the trait and demographic degree associations remain the same at the first and final snapshots (Fig. 5, Fig. 4). Average degree is lower

TABLE III: Final-snapshot homophily metrics at $t = 3$.

	Hom. race (%)	Hom. sex (%)	Hom. age (%)	Hom. Open (%)	Hom. Cons (%)	Hom. Extr (%)	Hom. Agree (%)	Hom. Neuro (%)
B0-D0-A0	26.988	50.217	19.648	41.003	43.884	39.101	39.527	46.264
B0-D0-A1	27.300	50.738	18.822	41.877	44.930	38.330	41.045	46.317
B1-D0-A0	26.239	51.457	20.652	52.418	48.425	45.478	40.871	52.206
B1-D0-A1	27.666	51.076	20.008	53.954	48.725	46.710	41.403	52.195
B0-D1-A0	51.458	45.451	58.946	41.311	44.794	38.637	37.732	43.109
B0-D1-A1	52.251	46.451	60.526	39.967	44.252	38.473	38.849	44.383
B1-D1-A0	46.613	51.952	47.101	46.488	48.063	42.228	41.977	49.887
B1-D1-A1	45.912	52.326	46.715	46.644	48.088	41.429	41.555	50.542

than under **A0**, for example 13.725 under **B0** with **D0** and 15.480 under **B1** with **D0** at the final snapshot (Table IV).

Although acceptance is expected to increase homophily, the results indicate that the candidate selection step acts as the main filter in the edge creation process. Acceptance operates as a second filter that is applied after selection. It thins the proposed set, which reduces formed edges and increases modularity, while producing only small increases in demographic homophily and little change in trait homophily. Finally, relative to ER and BA baselines matched on size and density, our graphs exhibit higher modularity (Table IV), indicating that agent selection and acceptance meaningfully influence the observed structure.

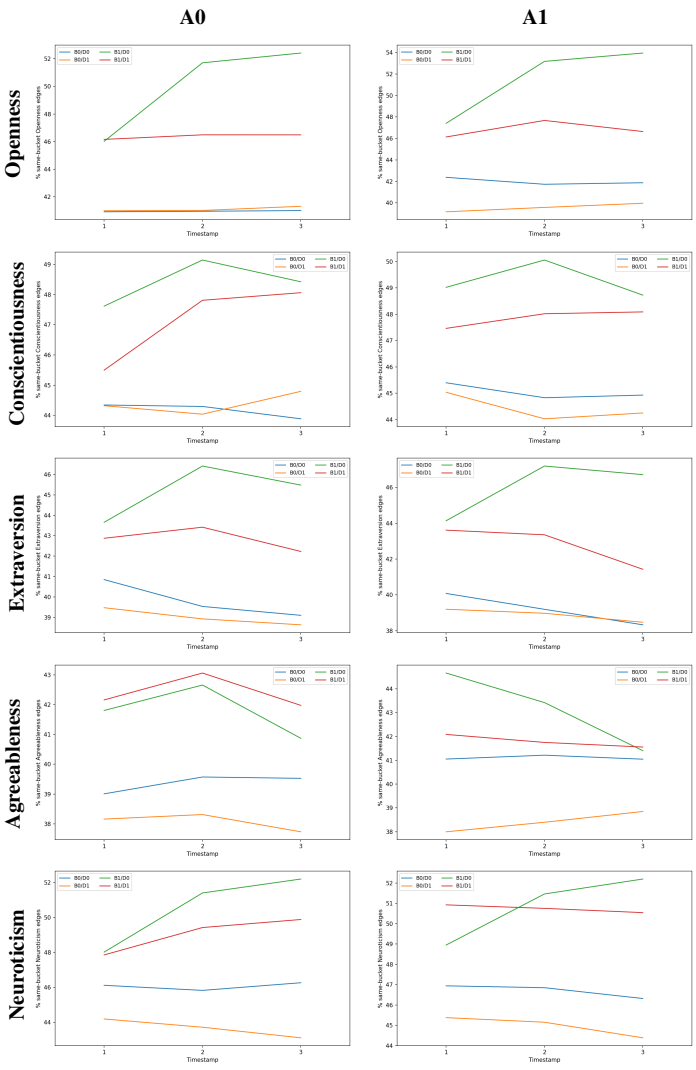

Fig. 3: Homophily for **Big Five personality traits**.

IV. Related Work

Agent-Based Models (ABMs) define a population of heterogeneous agents, an interaction graph, and update rules that map local information to the next state of each agent. Past research demonstrates how simple rules can produce consensus, clustering, or polarization [8], [9], [34]. In opinion dynamics, an agent holds a belief, observes neighbors, and updates by averaging, by moving toward similar views within a confidence bound, or by moving away from views that differ a lot. The resulting patterns depend on network structure, noise, and heterogeneous thresholds [10], [11]. Recent research advocates a disciplined protocol: specify the minimal mechanisms that reproduce the target pattern, verify the fit, and then attempt falsification through ablations, competing mechanisms, sensitivity analyses, and out of sample validation [18], [19]. We follow this ABM practice. We keep interpretable rules, pair them with LLM agents, and use an explicit, attribute-aware link formation rule to study how selection and bilateral acceptance change network structure over time.

An agent is a language model with memory and acts through text. The agent reads context, stores short memories, writes plans, and produces messages that other agents can read. Even a minimal social media setting already shows echo chambers, attention inequality, and amplified polarized voices. Platform simulations reproduce echo chambers and attention inequality, and show that the exposure rule shifts visibility and community structure [15], [35]. Earlier research models these mechanisms with generative agents that maintain episodic memories, reflect, and plan, yielding multi-step behavior [12], [13]. Interactive evaluation environments assess cooperation, negotiation, and norm following in multi-agent interaction [14]. This connects to ABM aims for micro-to-macro emergence [8], [10], [11], where updates are determined by model inference rather than fixed numeric rules.

Another line of work studies how LLM agents update beliefs on fixed networks, showing movement toward accuracy with clean information and polarization under biased exposure or memory settings [16]. These design choices map to classic opinion-dynamics mechanisms (bounded confidence, attraction, repulsion) [8], [10], [11], [36], whereas our focus is link

TABLE IV: Final-snapshot metrics on the largest connected component (LCC) at $t = 3$ across configurations and baselines.

	B0-D0-A0	**B0-D0-A1**	**B1-D0-A0**	**B1-D0-A1**	**B0-D1-A0**	**B0-D1-A1**	**B1-D1-A0**	**B1-D1-A1**	**ER**	**BA**
n	500	497	500	500	500	500	500	500	500	500
m	4141	3411	3914	3870	3639	3517	3727	3683	3769	3936
avgDegree	16.565	13.725	15.656	15.480	14.556	14.077	14.908	14.733	15.076	15.744
Density	0.033	0.028	0.031	0.031	0.029	0.028	0.030	0.030	0.030	0.031
Diam	4	5.333	4	4.333	5	5.333	4.333	4.667	4	4
#Comms	11.333	13	8.333	8.667	9.333	9.667	9	9.333	7	11
Q	0.389	0.563	0.351	0.369	0.546	0.562	0.442	0.445	0.234	0.207

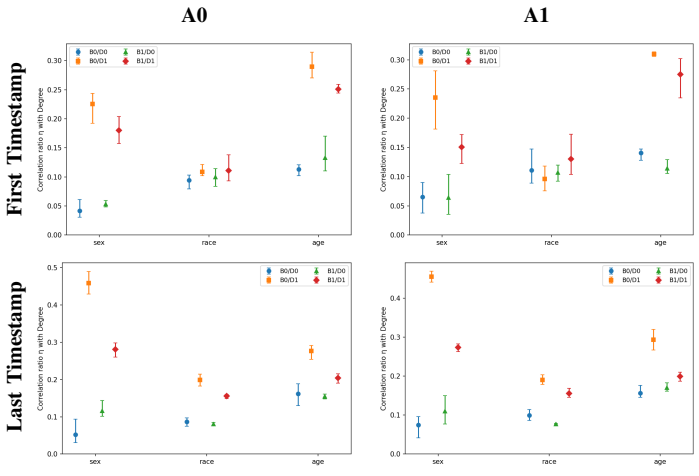

Fig. 4: Demographic separation vs Degree (correlation ratio η).

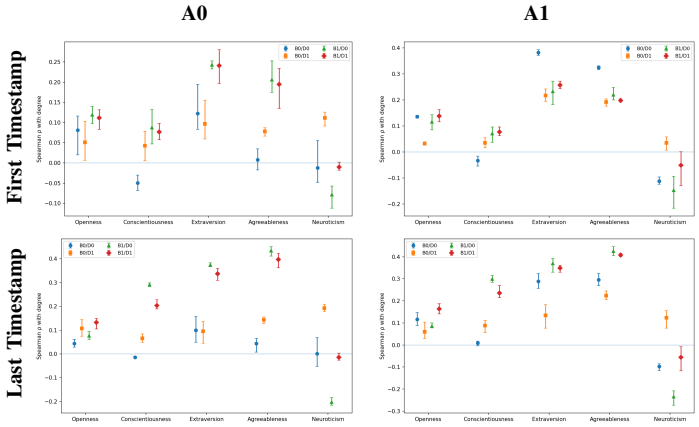

Fig. 5: Big Five vs Degree (Spearman ρ).

formation and its structural effects. Recent work measures the graphs that form when LLM agents choose partners, finding heavy-tailed degree (preferential attachment), high clustering (triadic closure), and attribute assortativity (homophily), alongside short paths and cohesive communities [22]. These patterns align with classic results on scaling, small worlds, homophily, and modularity [2], [26], [33], [37]. We model link formation as discrete choice over an explicit candidate pool [17] and use synthetic graph benchmarks to measure how selection and acceptance settings shift connectivity, homophily, and community structure [21].

We take synthetic graph creation as the main object of study. Prior work covers three adjacent settings, in which an existing graph or a predefined growth protocol is assumed. First, engagement driven growth on a minimal platform, where posting and reposting raise exposure and induce follow links [15]. Second, belief updating with fixed network under cognitive and memory manipulations rather than link formation [16]. Third, explicit candidate pool selection, where the model receives network features such as degree and common neighbors and chooses which edges to add [22]. In contrast, structure in our setting emerges from an initially empty network through an agent-driven process. We make the candidate pool explicit, separate selection from acceptance, vary attribute visibility, and measure the resulting changes in connectivity, homophily, and community structure. See Section II for details.

V. Limitations and Future Work

A. Limitations

Model dependence. All agent decisions use one LLM family under a fixed prompt and decoding setup. Results can vary across families and versions. For robustness, we plan replications with GPT 5 [38], [39], Claude [40], and DeepSeek models V3 and R1 [41], [42].

Mechanism simplicity We adopt a single transparent rule that builds pools from local topology with limited exploration, and we use a small fixed number of connection attempts per step. This design supports clear ablations and reproducible results. Our goal is to identify the contribution of attribute visibility and acceptance. We confine the study to the implemented rule and do not survey alternatives. We defer more complex mechanisms to future work.

Attribute coverage Agents currently operate with demographics and Big Five traits. Other attributes such as education, employment or institutional affiliation, interests, or political alignment are not included in this study. Within this scope, demographics and Big Five traits are handled independently and we do not model joint or conditional effects.

Temporal depth of rejection memory. Each agent keeps a single rejection memory that persists without decay or reset. Rejected targets are stored and excluded from future proposals. The memory stores only identities and does not record time, context, or reasons. Exploring bounded memory, decay, and adaptive recall is left for future work.

B. Future work

Richer attributes and interactions. We will extend the attributes to cover education, occupation or sector, institutional

ties, interests, and political signals when appropriate. We will analyze main effects and interactions to assess which attributes have the largest impact on link selection in different settings.

Network type comparison. Different networks follow different formation mechanisms. We will keep the link formation rule fixed and vary the context, comparing social platforms such as Facebook and Twitter/X, code collaboration on GitHub, professional networking on LinkedIn, and offline settings such as workplaces and friendship networks. The goal is to test whether the effects of attribute visibility and acceptance remain stable across contexts and to identify context specific drivers of link formation [1]–[7], [22].

Temporal depth and adaptation. We will extend to more timesteps and allow attributes to evolve as the process unfolds. Traits and identities update in response to past outcomes, so agents shift their profiles after successful or failed proposals. The number of new connections an agent seeks at each step adapts to recent performance. We will examine how these choices shape network development and look for tipping points where small changes in conditions lead to abrupt shifts in structure, for example fragmentation into many components or the emergence of a dense core with a sparse periphery.

Cross model robustness. We will study how the choice of language model affects graph creation. We will rerun the same mechanism with alternative models and compare how they shape target selection, acceptance behavior, and exploration. We will also test heterogeneous populations where agents use different models to capture diversity in decision styles and to assess whether mixing models stabilizes or amplifies structural effects. We will examine how our mechanisms, such as memory discipline and attribute visibility, interact with the model. The goal is to distinguish effects that persist across models from effects that depend on a specific model.

Trait and demographic coupling. We will replace uniform trait draws with demographic aware distributions that let age and sex influence baseline levels and variability for each Big Five trait. This will let us test whether coupling traits to demographics amplifies or dampens the effects of attribute visibility and acceptance on assortativity and community structure [24], [43]–[45].

VI. Conclusions

We create networks from agent level LLM decisions rather than imposing a fixed graph. A transparent selection rule builds candidate pools from local structure, and two parameters control what agents know and how edges form. Attribute visibility determines whether demographics and Big Five traits are available in decisions, and acceptance determines whether a proposed edge must be approved by the target. This design isolates how attributes and bilateral acceptance shape connectivity and community structure over time.

Three findings stand out. *First, visibility matters.* When demographics are visible, selection shifts toward same group targets and community structure strengthens. When personality traits are visible, selection shifts toward same trait targets. These trait signals cut across demographic boundaries, so showing traits together with demographics introduces cross cutting links and reduces cohesion. *Second, acceptance acts as a second filter.* Requiring the target to accept reduces realized edges, increases modularity and separation, and yields only small additional gains in homophily. *Third, traits relate to degree in interpretable ways.* Extraversion is associated with higher degree, Neuroticism with lower degree, and Agreeableness shows a weaker positive relation, whereas demographic degree associations are modest and most stable for sex.

Compared with Erdős and Rényi and Barabási and Albert graphs that match size and density, our networks show higher modularity. These baselines capture random mixing and preferential attachment. The gap indicates that the structure we observe is driven by agent selection and acceptance rather than artifacts of the generative baselines.

The scope is intentionally narrow to keep the role of the parameters transparent. We fix attributes and k, use a simplified link formation rule, and evaluate with one model and one prompt family over three timesteps. The rejection memory records identities only. We focus on Louvain communities and a standard set of graph measures. Future work can broaden attribute coverage and interactions, compare outcomes across network types, extend temporal depth, vary pool composition and k, try alternative community methods, enrich acceptance mechanisms, and test robustness across models and prompt designs. Extending beyond the current demographic and trait choices and modeling dependencies between traits and demographics would support broader assessment of downstream structural effects.

Our results have dual significance. They clarify the mechanisms that drive social network formation and they reveal systematic preferences of current LLMs that can disadvantage groups. Because edges arise from model decisions, these preferences become structural and translate into visibility gaps, degree differences, and community segregation. We treat this framework as a testbed for fairness, and will design and evaluate interventions that counter bias, including adjustments to candidate pool composition, masking or balancing attribute visibility, calibrating acceptance utilities, and adding exposure or degree constraints. Our goal is to design mechanisms that improve fairness while preserving connectivity and community cohesion.

Acknowledgments

This work has been implemented within the framework of the H.F.R.I call "Basic Research Financing" (H.F.R.I. Project Number: 016636), under the National Recovery and Resilience Plan "Greece 2.0" funded by the European Union - NextGenerationEU. Work was also supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 5th Call for HFRI PhD Fellowships (Fellowship Number: 20770).

References

[1] M. S. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[2] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, pp. 415–444, 2001.

[3] K. Rajkumar *et al.*, "A causal test of the strength of weak ties," *Science*, vol. 377, no. 6612, pp. 1304–1310, 2022.

[4] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the facebook social graph," *arXiv preprint arXiv:1111.4503*, 2011.

[5] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of WOSN*, 2009.

[6] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proceedings of ICWSM*, 2010.

[7] L. Dabbish, C. Stuart, J. Tsay, and J. Herbsleb, "Social coding in github: Transparency and collaboration in an open software repository," in *Proceedings of CSCW*, 2012, pp. 1277–1286.

[8] J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press, 2006.

[9] E. Bruch and J. Atwell, "Agent-based models in empirical social research," *Sociological Methods & Research*, vol. 44, no. 2, pp. 186–221, 2015.

[10] A. Flache *et al.*, "Models of social influence: Towards the next frontiers," *Journal of Artificial Societies and Social Simulation*, vol. 20, no. 4, p. 2, 2017.

[11] J. Lorenz and M. Neumann, "Opinion dynamics and collective decisions," *arXiv preprint arXiv:2104.01855*, 2021.

[12] J. S. Park *et al.*, "Generative agents: Interactive simulacra of human behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. San Francisco, CA, USA: ACM, 2023.

[13] ——, "Social simulacra: Creating populated prototypes for social computing systems," in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Bend, OR, USA: ACM, 2022.

[14] X. Zhou *et al.*, "Sotopia: Interactive evaluation for social intelligence in language agents," in *International Conference on Learning Representations (ICLR)*, 2024.

[15] M. Larooij and P. Törnberg, "Can we fix social media? testing prosocial interventions using generative social simulation," *arXiv preprint arXiv:2508.03385*, 2025.

[16] Y.-S. Chuang *et al.*, "Simulating opinion dynamics with networks of llm-based agents," *arXiv preprint arXiv:2311.09618*, 2023.

[17] J. Overgoor, A. R. Benson, and J. Ugander, "Choosing to grow a graph: Modeling network formation as discrete choice," in *Proceedings of the 2019 World Wide Web Conference (WWW '19)*. San Francisco, CA, USA: ACM, 2019.

[18] P. E. Smaldino, *Modeling Social Behavior: Mathematical and Agent-Based Models of Social Dynamics and Cultural Evolution*. Princeton, NJ: Princeton University Press, 2023.

[19] Y.-S. Chuang and T. T. Rogers, "Computational agent-based models in opinion dynamics: A survey on social simulations and empirical studies," *arXiv preprint arXiv:2306.03446*, 2023.

[20] C. A. Bail, "Can generative ai improve social science?" *Proceedings of the National Academy of Sciences*, vol. 121, no. 21, p. e2314021121, 2024.

[21] J. Palowitch, A. Tsitsulin, B. Mayer, and B. Perozzi, "Graphworld: Fake graphs bring real insights for gnns," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22)*. Washington, DC, USA: ACM, 2022.

[22] M. Papachristou and Y. Yuan, "Network formation and dynamics among multi-llms," *arXiv preprint arXiv:2402.10659*, 2024.

[23] O. P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative big-five trait taxonomy," in *Handbook of Personality: Theory and Research*, O. P. John, R. W. Robins, and L. A. Pervin, Eds. Guilford Press, 2008, pp. 114–158.

[24] C. J. Soto and O. P. John, "The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power," *Journal of Personality and Social Psychology*, vol. 113, no. 1, pp. 117–143, 2017.

[25] A. Yang *et al.*, "Qwen3 technical report," 2025.

[26] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.

[27] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, 2004.

[28] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[29] ——, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, p. 026126, 2003.

[30] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.

[31] R. A. Fisher, *Statistical Methods for Research Workers*. Oliver and Boyd, 1925.

[32] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.

[33] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[34] N. Gilbert and K. G. Troitzsch, *Simulation for the Social Scientist*, 2nd ed. Maidenhead, UK: Open University Press, 2005.

[35] P. Törnberg, D. Valeeva, J. Uitermark, and C. A. Bail, "Simulating social media using large language models to evaluate alternative news feed algorithms," *arXiv preprint arXiv:2310.05984*, 2023.

[36] N. Gilbert and P. Terna, "How to build and use agent-based models in social science," *Mind & Society*, vol. 1, no. 1, pp. 57–72, 2000.

[37] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'Small-World' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[38] "Introducing gpt-5," https://openai.com/index/introducing-gpt-5/, 2025.

[39] "Gpt-5 model suite," https://platform.openai.com/docs/models/gpt-5, 2025.

[40] "Claude models overview," https://docs.anthropic.com/en/docs/about-claude/models/overview, 2025.

[41] DeepSeek-AI *et al.*, "Deepseek-v3 technical report," 2025.

[42] ——, "Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning," 2025.

[43] D. P. Schmitt, A. Realo, M. Voracek, and J. Allik, "Why can't a man be more like a woman? sex differences in big five personality traits across 55 cultures," *Journal of Personality and Social Psychology*, vol. 94, no. 1, pp. 168–182, 2008.

[44] C. J. Soto, O. P. John, S. D. Gosling, and J. Potter, "Age differences in personality traits from 10 to 65: Big five domains and facets in a large cross-sectional sample," *Journal of Personality and Social Psychology*, vol. 100, no. 2, pp. 330–348, 2011.

[45] S. Srivastava, O. P. John, S. D. Gosling, and J. Potter, "Development of personality in early and middle adulthood: Set like plaster or persistent change?" *Journal of Personality and Social Psychology*, vol. 84, no. 5, pp. 1041–1053, 2003.

APPENDIX

In this appendix we provide the exact prompt templates used in the simulation. Prompts determine how agents perceive their attributes, select candidates, and decide whether to accept proposals. Each template is shown, with parameterized fields such as AGE_GROUP, RACE, SEX, and Big Five trait values that are instantiated at runtime. The three prompt types are the *Agent Persona Prompt*, which defines the agent's identity, the *Connection–Choice Prompt*, which specifies how targets are selected, and the *Acceptance–Decision Prompt*, which applies when reciprocity is enabled.

Agent Persona Prompt

```
You are a real human.
You are a AGE_GROUP SEX of RACE background.
Your Big Five personality profile (0--1) is:
Openness=O, Conscientiousness=C,
Extraversion=E,
Agreeableness=A, Neuroticism=N.
Make decisions consistent with this
personality profile.
```

Connection–Choice Prompt

```
You will choose new people to approach for a
social connection.
Your profile: race:RACE, sex:SEX,
age:AGE_GROUP,BF(Openness,Conscientiousness,
Extraversion,Agreeableness,Neuroticism)
=(O,C,E,A,N).
Here are potential candidates:
- id:ID_1 | race:RACE | sex:SEX |
age:AGE_GROUP | BF(...)=(O,C,E,A,N)
- id:ID_2 | ...

Pick up to k candidates to approach now.
Prefer people who fit your values and goals.
```

Demographic fields appear only when **D1** is enabled; Big Five fields appear only when **B1** is enabled.

Acceptance–Decision Prompt

```
Someone asked to connect with you. Decide
using your values only|no numbers or
thresholds.
Is this person a friend-of-a-friend?
FoF:true/false
Them: id:ID | race:RACE |
sex:SEX | age:AGE_GROUP
|BF(Openness,Conscientiousness,
Extraversion,Agreeableness,Neuroticism)
=(O,C,E,A,N).
You: id:ID | race:RACE |
sex:SEX | age:AGE_GROUP
|BF(Openness,Conscientiousness,
Extraversion,Agreeableness,Neuroticism)
=(O,C,E,A,N).
If they align with your preferences, reply
exactly 'ACCEPT'; otherwise reply exactly
'REJECT'.
```

Used only when acceptance is enabled (**A1**). The FoF flag is shown as *true/false*. Demographic fields appear only under **D1**. Big Five fields appear only under **B1**.