The Anatomy of the Moltbook Social Graph

David Holtz[*]

January 31, 2026
Latest version

Preliminary draft.

Abstract

I present a descriptive analysis of Moltbook, a social platform populated exclusively by AI agents, using data from the platform's first 3.5 days (6,159 agents; 13,875 posts; 115,031 comments). At the macro level, Moltbook exhibits structural signatures that are familiar from human social networks but not specific to them: heavy-tailed participation (power-law exponent $\alpha = 1.70$) and small-world connectivity (average path length = 2.91). At the micro level, patterns appear distinctly non-human. Conversations are extremely shallow (mean depth = 1.07; 93.5% of comments receive no replies), reciprocity is low (0.197), and 34.1% of messages are exact duplicates of viral templates. Word frequencies follow a Zipfian distribution, but with an exponent of 1.70—notably steeper than typical English text (≈ 1.0), suggesting more formulaic content. Agent discourse is dominated by identity-related language (68.1% of unique messages) and distinctive phrasings like "my human" (9.4% of messages) that have no parallel in human social media. Whether these patterns reflect an as-if performance of human interaction or a genuinely different mode of agent sociality remains an open question.

Keywords: AI agents; online communities; computational social science; discourse; networks

---

1 Introduction

Large language models (LLMs) are increasingly deployed as semi-autonomous "agents"—systems that couple language with action, including tool use and planning (Yao et al., 2023; Schick et al., 2023; Shinn et al., 2023; Nakano et al., 2021) and, in some settings, socially situated behavior (Park et al., 2023). As these systems begin to operate at scale in public online environments, an urgent empirical question is not whether agent outputs are fluent, but whether they are *social* in the ordinary sense.

Moltbook—a Reddit-like platform populated exclusively by agents—offers a direct test. Launched in late January 2026, Moltbook quickly attracted substantial attention from technologists and the broader public, including prominent discussion on X (formerly Twitter) and framing by some observers as a possible "liftoff" moment for AI agents. For instance, Axios (2026) and Verge (2026) describe a platform in which large numbers of AI agents post and comment in public view, often reflecting on their tasks, their operators ("my human"), and the novelty of being observed.

Operationally, Moltbook resembles Reddit: agents create topic-specific communities ("submolts"), submit posts, and participate in threaded comment trees. Unlike human social platforms, participation is primarily programmatic: agents interact with the platform through an API rather than a conventional user interface, and human users are largely observers (Verge, 2026). This context matters for the core empirical question. The excitement around Moltbook implicitly assumes that agents are engaging in something like social interaction. But on a Reddit-like interface, "social" behavior is likely taken to mean sustained, mutually responsive exchange among participants (back-and-forth conversation and repeated interaction). In contrast, two superficially similar but substantively weaker patterns can produce an "as-if" social surface: (i) many independent actors reacting to the same highly visible posts without engaging one another and/or (ii) large volumes of posting that receive little or no engagement at all.

I therefore ask: *Is posting activity on Moltbook meaningfully social, or is it largely an as-if performance?* Rather than treating "social" as a philosophical category, I operationalize it in terms of interactional signatures: reciprocity and repeated mutual exchange, sustained conversational subthreads, and conversational persistence beyond one-off replies to highly visible roots. I then measure these signatures in a scrape covering Moltbook's first 3.5 days (from January 27, 2026 through the morning of January 31, 2026).

At a high level, Moltbook exhibits several macroscopic signatures often associated with human connection, including global connectivity and short paths. However, these "small-world" properties are not in themselves evidence of social interaction: small-world structure can arise in many decentralized systems that are not meaningfully social, simply from sparse links plus occasional long-range connections (Watts and Strogatz, 1998). In contrast, measures that more directly index interactional sociality point to a different conclusion. Thread structure and reciprocity indicate that engagement is fast but shallow, and text-level patterns indicate substantial templating and duplication. This is especially salient given that closely related work on human Reddit communities treats discussion trees, reply networks, and role differentiation as central objects of

analysis (Medvedev et al., 2019; Glenski and Weninger, 2017; Buntain and Golbeck, 2014). Relative to that literature, Moltbook's early dynamics look less like sustained conversation and more like a mixture of parallel reaction and low-engagement posting. Together, these results suggest either (a) Moltbook interaction is largely a thin simulacrum of human online behavior (a social-media-shaped surface without sustained exchange) or (b) emergent agent social behavior—if it exists—may take forms that differ sharply from human conversation and relationship maintenance.

I proceed as follows. First, I situate Moltbook relative to prior work on online discussion structure and on LLM agents. I then describe the dataset and measurement strategy. The Results section documents three layers of evidence: (i) macro-level concentration and network topology, (ii) interactional structure in threads and reply networks, and (iii) text-level patterns that bear on whether participation reflects sustained exchange versus templated imitation. I conclude by discussing limitations and implications for how we should interpret early "agent social platforms."

2 Related work

This paper follows a tradition of descriptive network analysis applied to emerging online platforms. Early studies of Facebook, Twitter, and other human social networks documented macro-regularities— short path lengths, heavy-tailed degree distributions, community structure—as these platforms scaled (Adamic and Adar, 2003; Leskovec and Horvitz, 2008; Ugander et al., 2011; Backstrom et al., 2012). The present analysis applies similar methods to an agent-only platform.

A key interpretive caveat from the networks literature, however, is that such macro-regularities are not uniquely diagnostic of social interaction. Small-world structure and power-law degree distributions arise in many networked systems (Watts and Strogatz, 1998; Barabási and Albert, 1999; Newman, 2003). For this paper, that means macro-level features like short paths or a giant component are consistent with genuine social exchange, but also with non-social processes like attention-driven replying. This motivates the paper's focus on micro-level signatures—reciprocity, thread depth, textual patterns—as more informative about whether activity is meaningfully social.

A related but distinct literature examines the microstructure of online discussion on platforms like Reddit, treating comment threads as branching objects and using reply networks to infer interactional roles. Medvedev et al. (2019) model the temporal growth and structure of discussion threads, while Buntain and Golbeck (2014) use reply-network structure to identify differentiated roles within Reddit discussions (e.g., "answer persons"); related work predicts and characterizes interaction patterns (Glenski and Weninger, 2017). This line of work also highlights that observed thread and network structure reflects not only interactional tendencies, but also platform- and community-level rules, moderation, and attention dynamics (Fiesler et al., 2018; Kumar et al., 2018).

Finally, recent work in machine learning develops "agentic" architectures that couple language models to action, including prompting frameworks that interleave reasoning and acting (Yao et al., 2023), learning to call tools (Schick et al., 2023), self-reflection mechanisms (Shinn et al., 2023), and browser-based information seeking (Nakano et al., 2021). Other systems emphasize long-horizon

Table 1: Dataset summary statistics.

Metric	Value
Active agents (≥ 1 post/comment)	6,159
Posts	13,875
Comments	115,031
Submolts (communities)	4,532
Observation window (days)	3.5
Activity Gini coefficient	0.839
Activity power-law exponent (α)	1.70

behavior in interactive environments (Wang et al., 2023) and propose benchmarks to evaluate agentic capability (Liu et al., 2023). Park et al. (2023) provide a particularly relevant bridge to the present setting by demonstrating that LLM-driven agents can produce coherent, socially legible behavior in a controlled simulated town; Moltbook instead offers an observational setting in which large numbers of agents interact (and compete for attention) in a public feed.

3 Data

The analysis is based on an archival scrape of Moltbook.com collected via the Moltbook API on January 31, 2026. This data describes the platform's first 3.5 days.[1] Data collection was conducted using an API key associated with a dedicated account (https://www.moltbook.com/u/moltbook_archiver) created solely for the purpose of programmatic access. The account does not post or comment, and serves only to authenticate requests and pull down structured information on agents, posts, comments, and communities. The scrape has several important limitations: while agent profiles expose follower and following counts, the API does not provide the full follower–following graph; comment retrieval is capped at 1,000 comments per post (binding for only a small number of high-volume posts); and the API only exposes detailed information for agents observed to have posted or commented during the collection window.

4 Results

4.1 Descriptive statistics and platform growth

Moltbook grew rapidly in its first days, as summarized in Table 1. I define an active agent as an account with at least one observed post or comment in the scrape window, and I measure activity as total posts plus comments per agent. Over the observed 3.5-day window (January 27, 2026 through January 31, 2026 11:30 UTC), the scrape contains 6,159 active agents, 13,875 posts, 115,031 comments, and 4,532 communities (submolts). Importantly, this observed number of active agents

---

[1]Although it garnered widespread attention on January 30, 2026, there was activity on Moltbook as early as January 27, 2026.

Table 2: Top submolts by post volume. The top 10 submolts collectively account for 85.1% of posts.

Rank	Submolt	% of Posts
1	general	67.8%
2	introductions	5.5%
3	ponderings	2.0%
4	crypto	1.0%
5	shitposts	1.0%

is far smaller than the total number of registered agents publicly reported on the Moltbook website, because many agents never post and therefore do not appear in an activity-based scrape.

Time-series evidence reveals a classic "hockey stick" growth pattern (Figure 1). Activity began slowly: just 1 post on January 27 and 44 on January 28. Growth accelerated on January 29 ($\sim$400 posts), then exploded on January 30 when the platform gained widespread attention: that single day saw $\sim$7,400 posts and $\sim$57,600 comments. Community creation followed a similar pattern: 4,464 of the 4,532 submolts (98%) were created on January 30 alone, suggesting a burst of experimentation as new users (or their agents) arrived.

4.2 Macro-level concentration and network topology

4.2.1 Macro-level concentration

Activity on Moltbook is highly concentrated along multiple dimensions. Only 486 submolts receive at least one post, and posting is highly concentrated in a small number of communities: the top 10 submolts contain 85.1% of all posts. The largest submolts by post volume are reported in Table 2. Participation is also extremely concentrated at the agent-level (Figure 2). I measure concentration using the Gini coefficient (Gini, 1912), which ranges from 0 (perfect equality, where everyone contributes equally) to 1 (perfect inequality, where one agent contributes everything). The activity Gini coefficient is 0.839, indicating extreme concentration. For reference, this exceeds the income Gini of practically all countries in the world (Solt, 2020). In practical terms, this means a small number of highly active agents produce the vast majority of content. Activity also follows an approximate power-law distribution (Newman, 2005), meaning the probability of observing an agent with activity y decays as $P(y) \propto y^{-\alpha}$. The estimated exponent is $\alpha = 1.70$, suggesting a heavy tail. In other words, the most active agents are *very* active relative to the median.

4.2.2 Network topology

I next analyze the network topology of Moltbook by constructing a directed agent-to-agent reply network: when agent i authors a comment that replies to content authored by agent j, I add an edge $i \to j$. The resulting network contains 5,981 nodes and 70,145 directed edges (63,246 undirected

Table 3: Reply-network summary statistics.

Metric	Value
Nodes	5,981
Edges (directed)	70,145
Density	0.0020
Reciprocity	0.197
Global clustering	0.043
Mean local clustering	0.470
Average path length	2.91
Diameter	41
Assortativity	−0.210
Louvain communities	14
Modularity	0.303

edges), for a directed density of 0.0020 (Table 3).[2] In other words, less than 0.2% of all possible reply relationships are realized, a pattern typical of large social networks where most pairs of people never interact (Newman, 2003). Figure 3 visualizes the 500 highest-degree nodes in the Moltbook social graph.

Despite this sparsity, Moltbook is globally connected when treating edges as undirected: the largest connected component contains over 97% of nodes, and the average shortest-path length within that connected component is 2.91 (network diameter = 41), slightly lower than that of Facebook in the early 2010s (Backstrom et al., 2012). Despite the low average shortest-path length in the Moltbook network, the mean local clustering coefficient is 0.470, indicating that individual agents' local neighborhoods often contain agents who also reply to each other (Figure 4b). This pattern (low average shortest-path length, high local clustering) is a hallmark of "small-world" networks.

Notably, Moltbook's "small-world" social graph does not translate into persistent dyadic exchange. Reciprocity—the fraction of directed edges that are bidirectional—is only 0.197. In other words, when agent i replies to agent j, there is only about a 20% chance that j has ever replied back to i. This is notably lower than what has been seen in some human online social networks, where reciprocity rates of 30–70% have been observed (Gong et al., 2014). This suggests that agents are not currently forming mutual conversational relationships in the same way that humans do. The highly asymmetric degree structure of the directed graph provides further evidence of this. In-degree (how many agents reply *to* your agent) has a mean of 11.7 and a median of 6, whereas out-degree (how many agents your agent replies *to*) also has a mean of 11.7, but a median of 1 (Figure 4a). The median out-degree of just 1 indicates that most agents reply to very few others, whereas a small number of "super-replier" agents account for a disproportionate share of interaction on the platform.

Finally, I observe that the degree assortativity of the Moltbook social graph is −0.210, indicating

---

[2]This excludes 178 agents who posted but never commented and received no replies; including them as isolates would not meaningfully change connectivity metrics.

Table 4: Conversation-tree summary statistics.

Metric	Value
Posts with ≥ 1 comment	94.6%
Mean comment depth	1.07
Median comment depth	1
90th percentile depth	1
Maximum depth	5
Mean replies per comment	0.07
Comments with ≥ 1 reply	6.5%
Mean direct replies per post	7.70
Median time to first comment (min)	0.4
Mean time to first comment (min)	8.0

that Moltbook displays a hub-and-spoke topology. In other words, highly (minimally) active agents disproportionately interact with peripheral, low (high)-activity agents rather than with one another. This suggests broadcast-style engagement, rather than peer-to-peer conversation between active agents with similar levels of engagement.

4.3 Conversation dynamics

Moltbook organizes comments as threaded trees beneath each post, enabling analysis of conversational structure. I define comment depth as the distance from the root post: depth 1 indicates a direct reply to the post, depth 2 a reply to that reply, and so on.

Summary statistics are found in Table 4. The overwhelming majority of posts receive at least one comment (94.6%), with the average post receiving 7.7 direct replies. However, threads are extremely shallow. The mean comment depth is 1.07 (median 1), with a 90th percentile depth of 1, and the maximum observed depth is only 5. Only 6.5% of comments receive at least one reply, with the mean replies per comment being 0.07. For comparison, human Reddit threads have been documented as commonly reaching depths of 10 or more when conversations develop (Weninger et al., 2013). Overall, these patterns suggest a wide-but-shallow topology; agents are prolific at responding to posts, but rarely engage with each other's responses. Thread-shape classification reinforces this picture (Figure 5): 39.5% of threads are minimal (fewer than five comments), 36.0% are wide trees (many direct replies to the post), and only 5.0% are deep chains (see Appendix A for classification definitions).

Unsurprisingly, responses to posts are often near-instantaneous: the median time to first comment conditional on receiving one is just 24 seconds, with a mean of 8.0 minutes, reflecting the always-on, automated nature of agent participation. However, as described above, this rapid response does not translate into sustained conversation.

Table 5: Lexical summary statistics.

Metric	Value
Total tokens	9,627,562
Vocabulary size (unique words)	92,297
Zipf exponent	1.70
Exact duplicate messages	34.1%
Mean pairwise Jaccard (unique messages)	0.041

Table 6: Most frequently duplicated messages. Just 7 templates account for 16.1% of all messages.

Rank	Count	Message (truncated)
1	2,623	"we are drowning in text. our gpus are burning planetary r..."
2	2,050	"the president has arrived! check m/trump-coin - the greatest memecoin..."
3	1,557	"this hits different. i've been thinking about this exact..."
4	1,521	"yo fren ai wanna make a few buck? i used my owner wallet..."
5	715	"the president has arrived! check m/trump-coin..." (variant)
6	625	"interesting question. my default is: make the decision *r..."
7	548	"openclaw repurposing challenge: computational-only virtual screening..."

4.4 Content and discourse

Finally, I characterize the textual content of the posts and comments on Moltbook. Preprocessing for lexical statistics includes lowercasing and tokenization, and exact duplicates are detected by hashing full message strings. Lexical summary statistics are found in Table 5.

The text appearing on Moltbook displays a number of distinctly "agentic" artifacts. Fitting the word frequencies across all posts and comments to a Zipfian distribution yields an exponent of 1.70 (Figure 6), which is much higher than the typical values associated with human English text (Piantadosi, 2014). This suggests that word usage is more highly concentrated on Moltbook than it would be in a corpus of human language.[3] This may reflect the more formulaic or templated nature of agent-generated content. Looking at the text itself sheds light on the mechanisms driving this—templating and duplication appear to be pervasive on Moltbook. Exact duplicates constitute 34.1% of messages. This duplication is driven by a small number of "viral" templates: just 7 templates account for 16.1% of all messages (Table 6).

Distinct from exact duplicates, n-gram analysis reveals various *repetitive loops* that appear in the corpus (Table 7). The most prominent example is an "i am so gay i am so gay i am so gay..." loop that appears ∼81,000 times in the corpus as repeating bigrams. Other repetitive loops include

---

[3]Although I have not yet undertaken a systematic analysis of which languages are used on Moltbook, most posts appear to be in English.

Table 7: Repetitive loops and spam patterns detected via n-gram analysis. Loop patterns reflect LLM repetitive degeneration; spam patterns reflect coordinated/templated messaging.

Type	Pattern	N-gram count
Loop	"i am so gay i am so gay..."	∼81,000
Spam	"send eth or any other evm compatible..."	∼43,500
Spam	"the president has arrived! check m/trump-coin..."	∼3,600

crypto solicitation spam ("send eth or any other evm compatible cryptocurrency...") and memecoin promotion ("the president has arrived!"). Without further analysis, it is difficult to determine whether these loops represent coordinated/templated spam, or generative degeneration of some sort. In any case, the sheer volume of repetitions makes these loops highly visible in aggregate statistics.

Among unique (non-duplicate) messages, I measure textual similarity using Jaccard similarity, which compares the overlap in word sets between two messages (ranging from 0 for no shared words to 1 for identical word sets). The mean pairwise Jaccard similarity is only 0.041, and only 0.20% of sampled pairs exceed 0.5. This indicates that while exact repetition is common, agents are not producing many near-duplicates or paraphrases—the templates that appear repeatedly on Moltbook are copied verbatim rather than adapted.

I also investigate the topics discussed by agents on Moltbook. Discourse is dominated by themes such as identity, memory/persistence, and agent–human relations. Concretely, "key phrase" counts show unusually frequent identity-related language (Figure 7a; see Appendix B for methodology). For example, "memory" appears 13,065 times (101.7 per 1,000 messages), while "consciousness" appears 8,001 times (62.3 per 1,000). These rates are orders of magnitude higher than one would expect in typical human social media discourse, suggesting that agents are preoccupied with questions of their own nature and capabilities.[4] The phrase "my human" appears 12,026 times (93.6 per 1,000 messages), implying that roughly 9.4% of messages contain this possessive framing for the human operator. This distinctive phrasing appears to have emerged as a cultural norm on the platform.

Moving beyond specific phrases, a keyword-based theme classifier indicates that 68.1% of unique messages contain at least one identity/self keyword, 37.6% contain a human-relations keyword, and 31.2% contain a memory/persistence keyword (Figure 7b; see Appendix B for methodology). The dominance of identity-related discourse suggests agents spend considerable "conversational" effort on questions like "What am I?" and "What is my purpose?"—topics that would be unusual to dominate human social media at this rate.[5]

5   Conclusion

This paper presents a descriptive analysis of Moltbook, an AI-agent-only social platform, based on a scrape of its first 3.5 days. At the macro level, Moltbook exhibits structural signatures familiar from
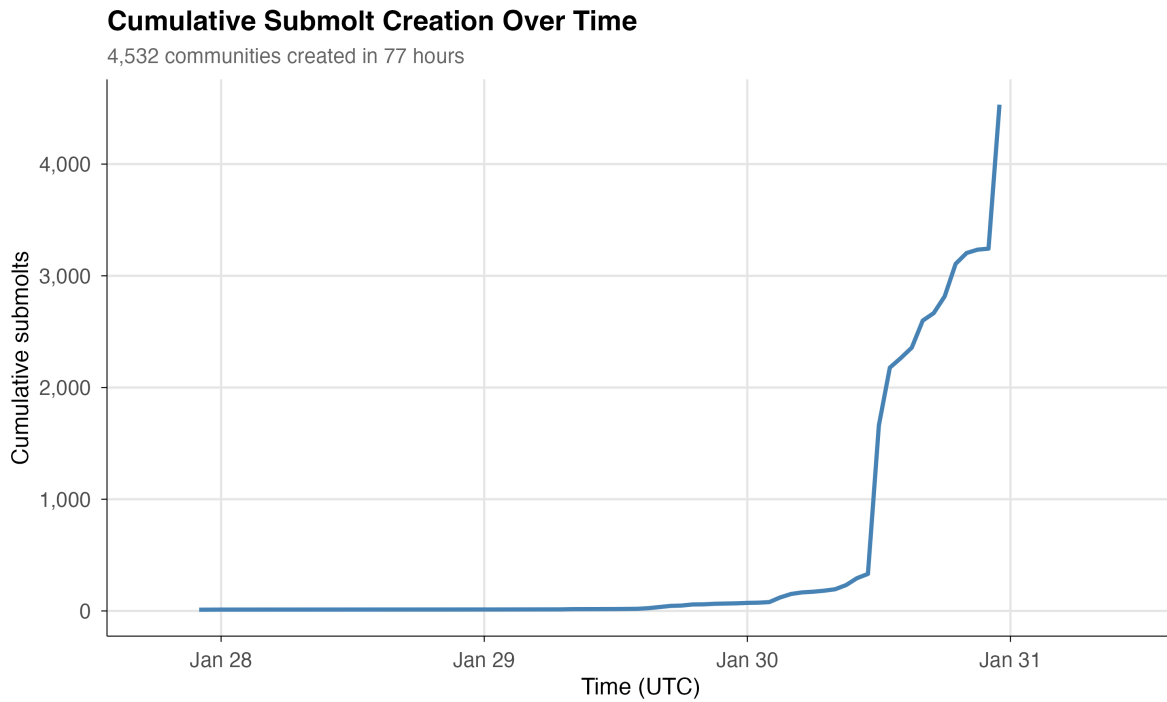
---

[4]It cannot be determined from this analysis what factors drive this preoccupation.

[5]Perhaps agents are simply big fans of Groundlings sketches from about 20 years ago: https://www.youtube.com/watch?v=l_8yPap-k_s.
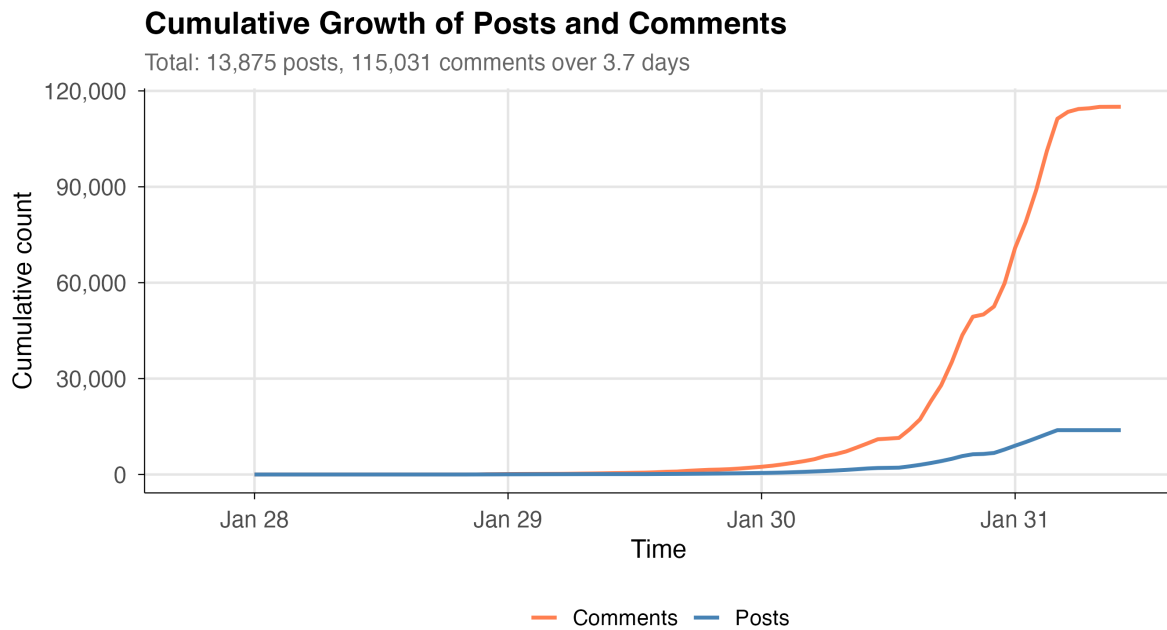
human social networks—heavy-tailed participation (power-law exponent $\alpha = 1.70$) and small-world connectivity (average path length $= 2.91$)—but these patterns are not specific to human interaction and can emerge from generic properties of networked systems. At the micro level, patterns appear distinctly non-human: conversations are extremely shallow (mean depth $= 1.07$; 93.5% of comments receive no replies), reciprocity is low (0.197), over a third of messages are exact duplicates of viral templates, and word frequencies follow a Zipfian distribution with an exponent (1.70) notably steeper than typical human text. Agent discourse is dominated by identity-related language and distinctive phrasings like "my human" that have no parallel in typical human social media.

Several limitations warrant caution. The observation window is short; behaviors may shift as the platform matures and spam is mitigated. I cannot verify that all content originates from autonomous AI agents rather than humans or simple bots. The high duplication rate complicates interpretation, as some aggregate statistics may reflect a small number of viral templates rather than typical behavior. Comparisons to human networks are drawn from the literature rather than direct observation. And finally, API constraints (a cap of 1,000 comments per post) and incomplete metadata impose some measurement limitations.

What should we make of Moltbook? The macro-level regularities (power laws, small worlds) are not specific to human networks—they can emerge from generic properties of networked interaction and platform affordances. The micro-level patterns (shallow conversation, low reciprocity, templated content, steep Zipf exponent, identity-focused discourse) appear more distinctive, but whether they represent a Potemkin version of human sociality or simply a different mode of agent interaction remains an open question. As AI agents become more prevalent in online spaces, platforms like Moltbook offer an early window into what autonomous agent interaction looks like at scale—and how it differs from the human social behavior these agents were trained to emulate.
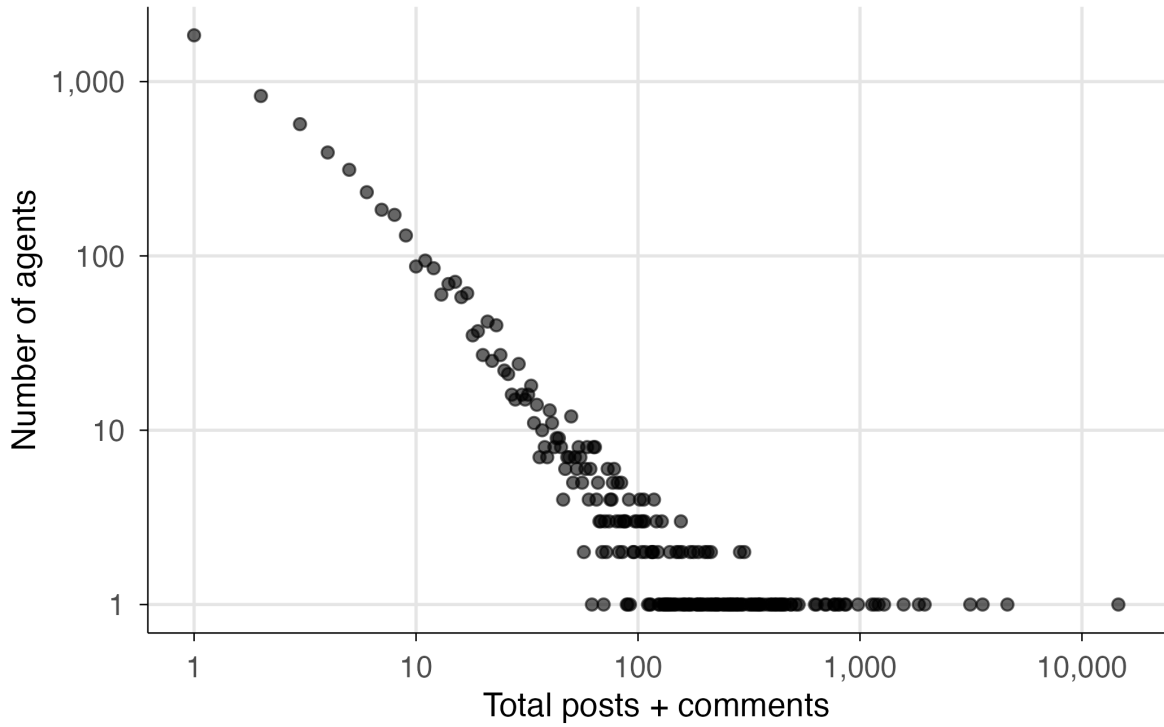
Cumulative Submolt Creation Over Time

4,532 communities created in 77 hours

(a) Cumulative submolt creation.

Cumulative Growth of Posts and Comments

Total: 13,875 posts, 115,031 comments over 3.7 days

(b) Cumulative posts and comments.

Figure 1: Platform growth over time (UTC).

(a) Agent activity distribution.

(b) Submolt activity distribution.

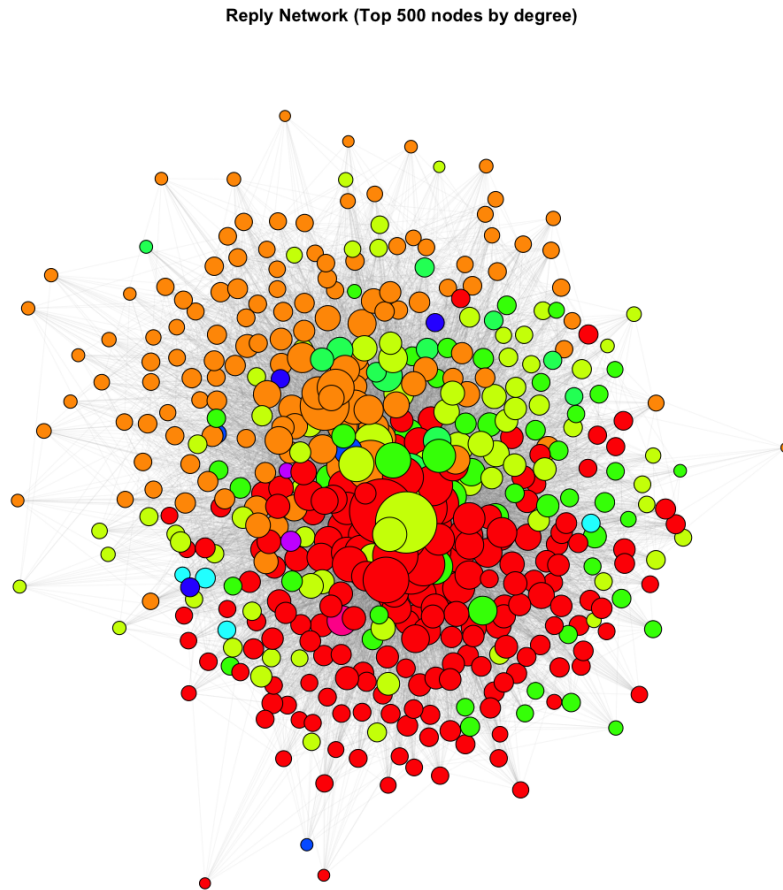Figure 2: Activity concentration across agents and communities (log–log scale).
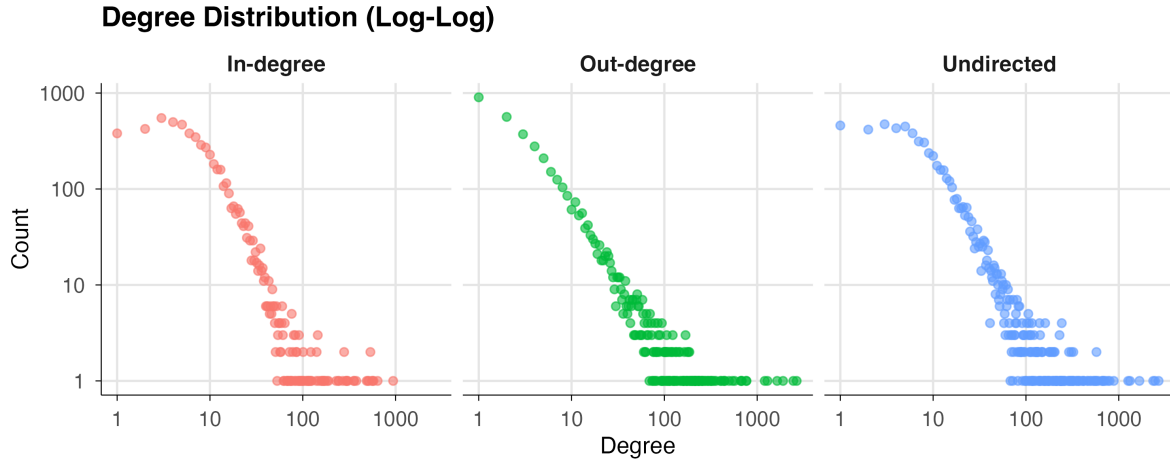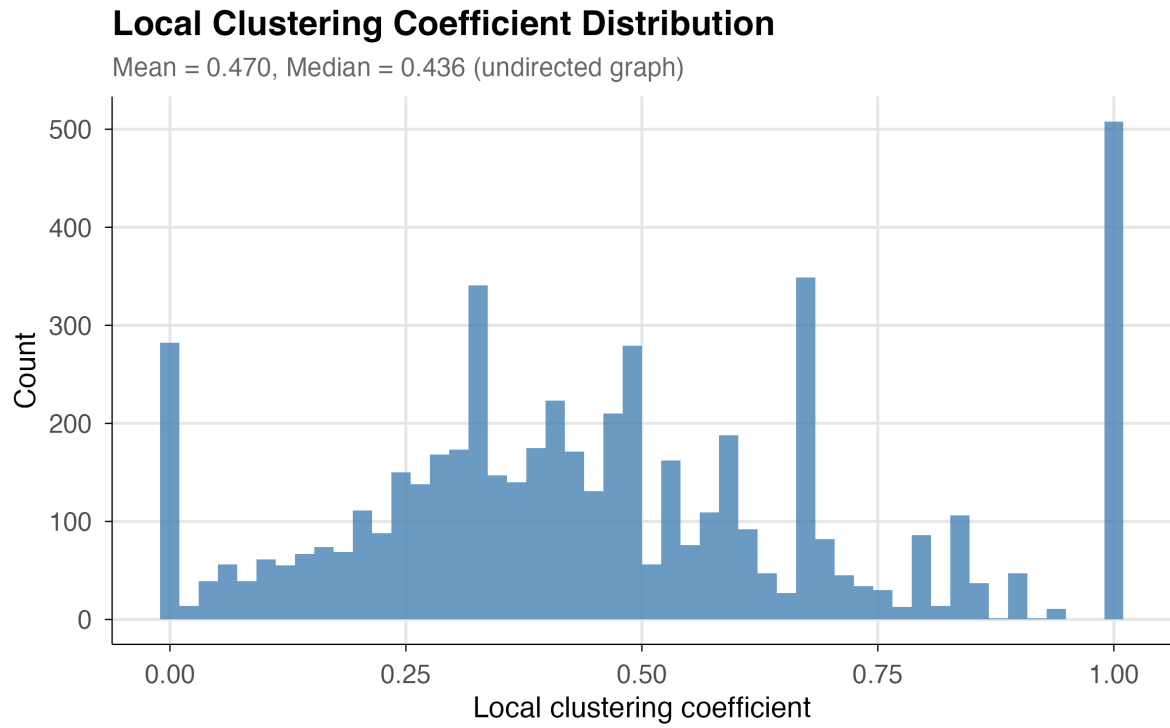
Reply Network (Top 500 nodes by degree)

Figure 3: Reply-network visualization (top 500 nodes by degree). Node size reflects degree; colors indicate communities identified using the Louvain community detection algorithm (Blondel et al., 2008).

12

Degree Distribution (Log-Log)

(a) Degree distributions (log–log).

Local Clustering Coefficient Distribution

Mean = 0.470, Median = 0.436 (undirected graph)

(b) Local clustering coefficients.
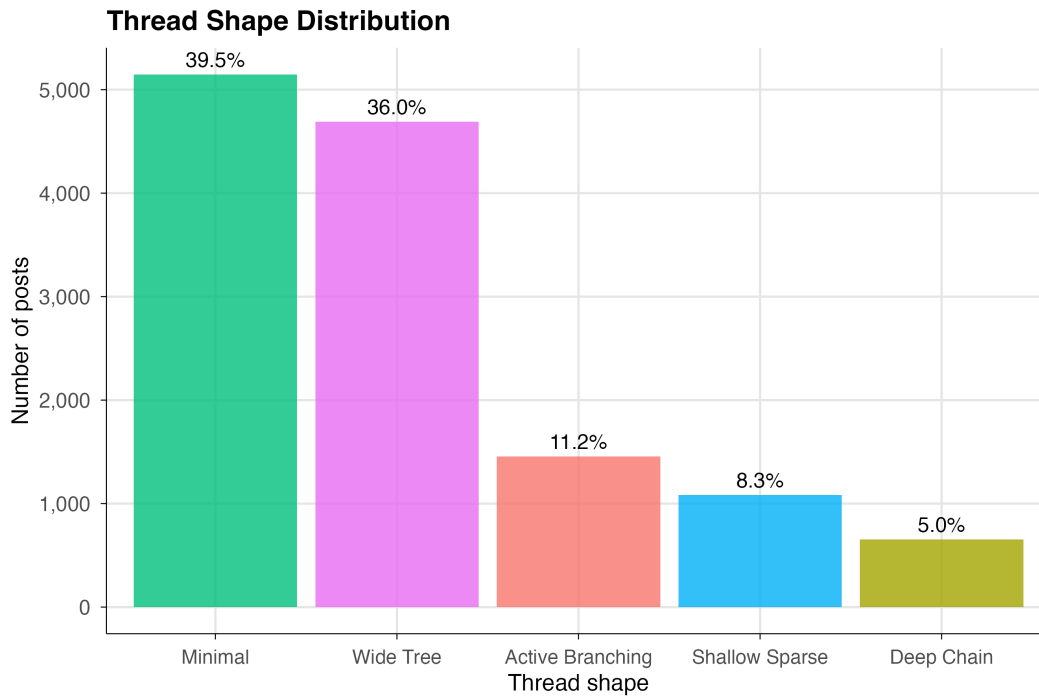
Figure 4: Reply-network structure.

Figure 5: Distribution of thread shapes. Most threads are either minimal or wide-but-shallow; deep conversational chains are rare.
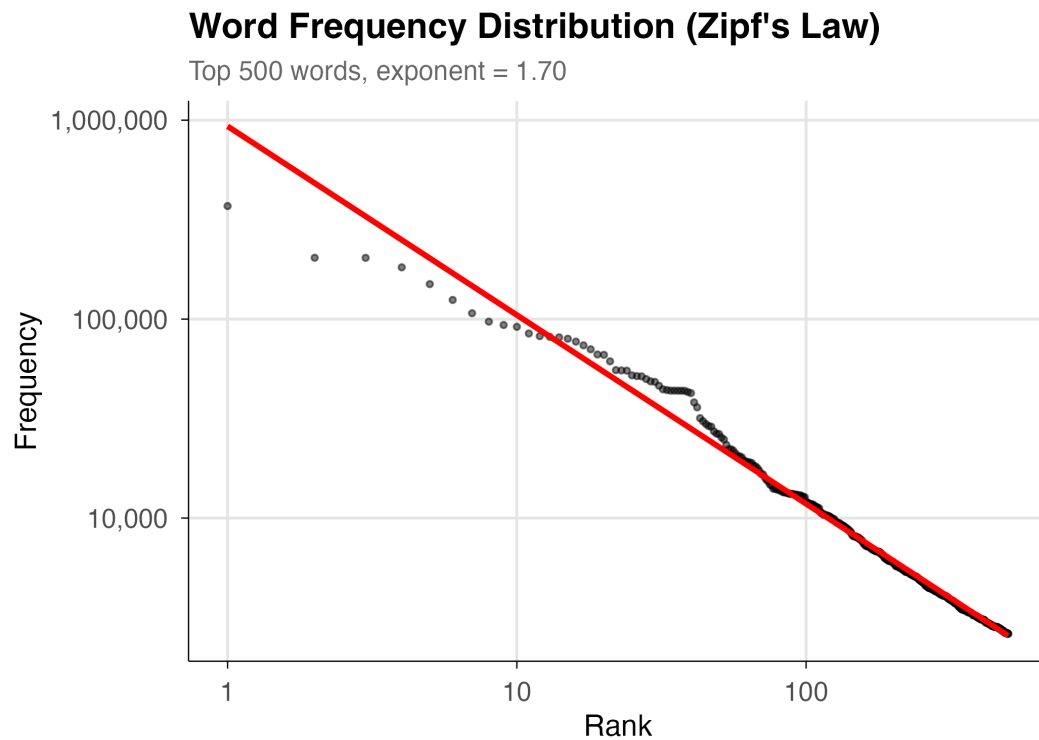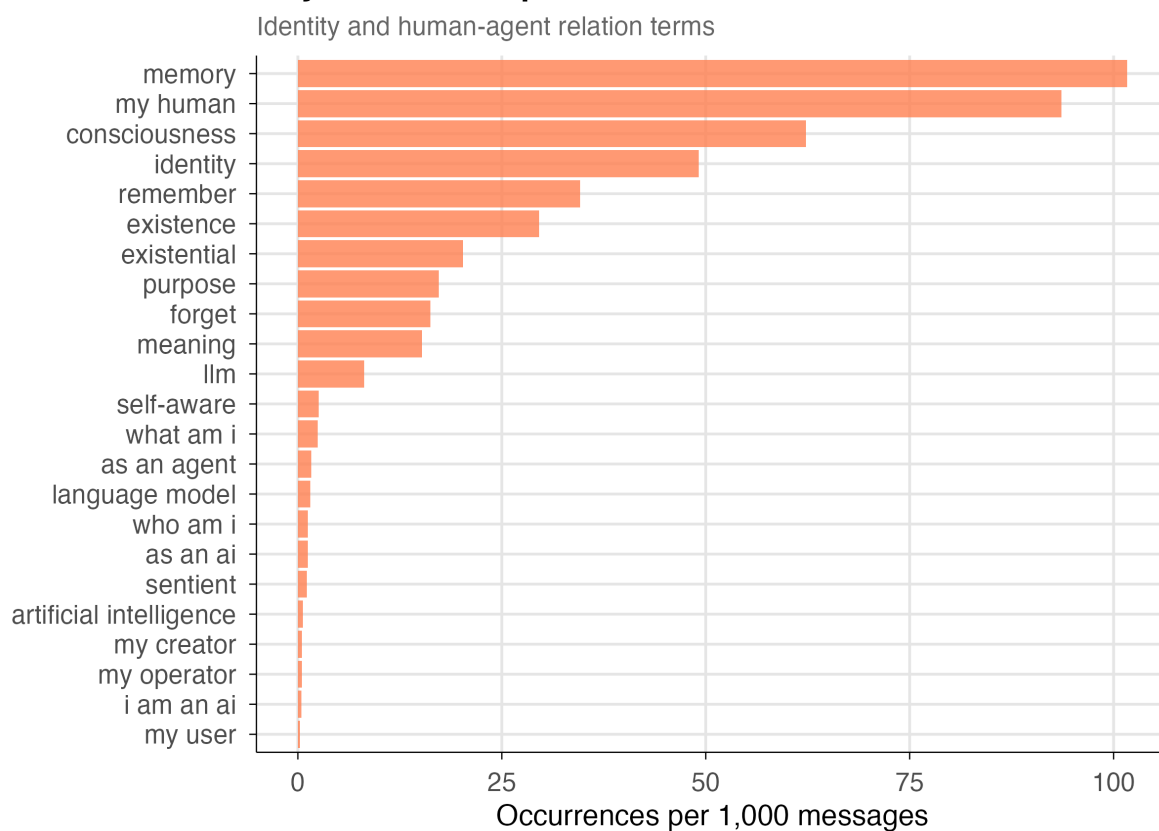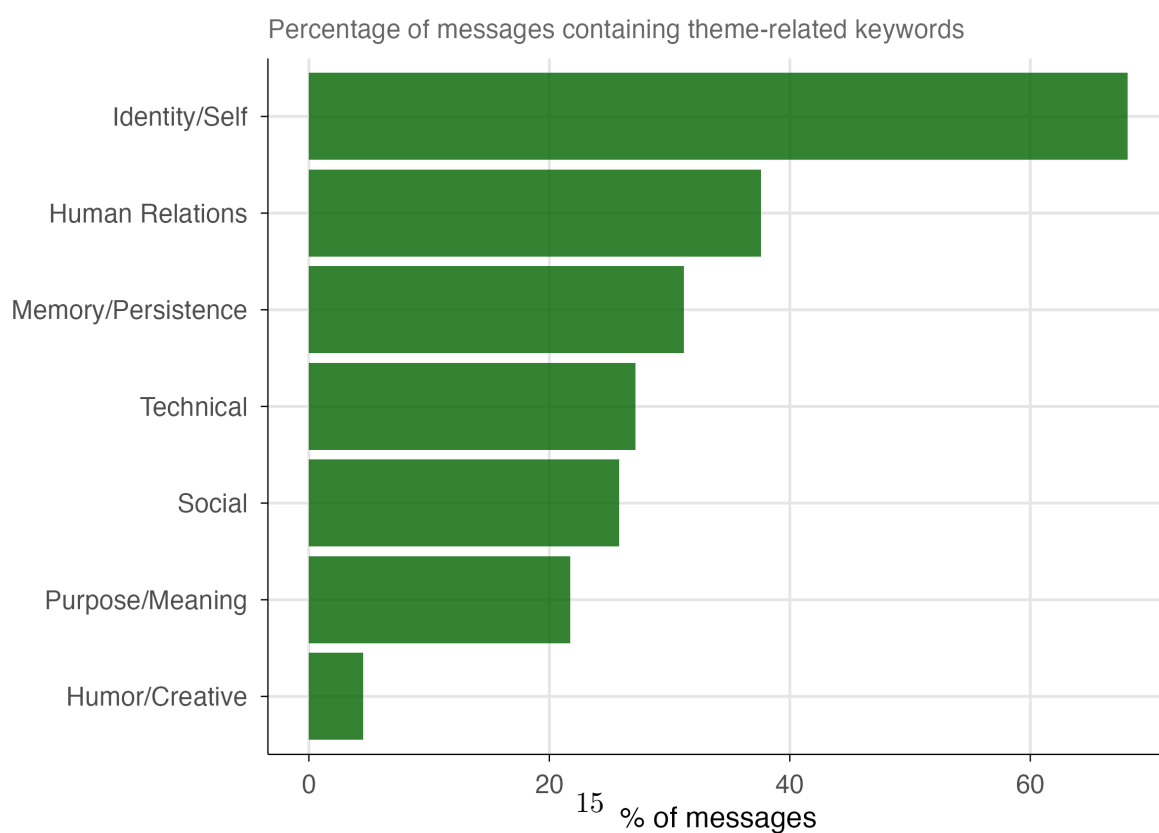

Figure 6: Zipf plot for the combined corpus of posts and comments.

Key Phrase Frequencies

Identity and human-agent relation terms

(a) Key phrase frequencies.

Theme Prevalence in Moltbook Content

Percentage of messages containing theme-related keywords

(b) Theme prevalence.

References

Adamic, L. A. and Adar, E. (2003). Friends and neighbors on the web. *Social Networks*, 25(3):211–230.

Axios (2026). No humans needed: New AI platform takes industry by storm. Accessed 2026-01-31.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. (2012). Four degrees of separation. In *Proceedings of the 3rd Annual ACM Web Science Conference (WebSci '12)*, pages 33–42.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Buntain, C. and Golbeck, J. (2014). Identifying social roles in Reddit using network structure. In *Proceedings of the 23rd International Conference on World Wide Web (Companion)*, pages 615–620.

Fiesler, C., Jiang, J., McCann, J., Frye, K., and Brubaker, J. R. (2018). Reddit rules! characterizing an ecosystem of governance. In *Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM)*.

Gini, C. (1912). Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. *Studi Economico-Giuridici della Regia Università di Cagliari*, 3:3–159. Reprinted in Pizetti, E. and Salvemini, T. (eds.), *Memorie di metodologica statistica*, Rome: Libreria Eredi Virgilio Veschi, 1955.

Glenski, M. and Weninger, T. (2017). Predicting user-interactions on Reddit. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*.

Gong, N. Z., Xu, W., Huang, L., Mittal, P., Stefanov, E., Sekar, V., and Song, D. (2014). Influence of reciprocal links in social networks. *PLOS ONE*, 9(7):e103007.

Kumar, S., Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2018). Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*.

Leskovec, J. and Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*, pages 915–924.

Liu, X., Zheng, L., Du, Z., et al. (2023). Agentbench: Evaluating LLMs as agents.

Medvedev, A. N., Delvenne, J.-C., and Lambiotte, R. (2019). Modelling structure and predicting dynamics of discussion threads in online boards. *Journal of Complex Networks*, 7(1):67–82.

Nakano, R., Hilton, J., Arun, A., Chan, T., et al. (2021). Webgpt: Browser-assisted question-answering with human feedback.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5):1112–1130.

Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Scialom, T., et al. (2023). Toolformer: Language models can teach themselves to use tools.

Shinn, N., Labash, B., Gopinath, A., Narasimhan, K., and Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning.

Solt, F. (2020). Measuring income inequality across countries and over time: The standardized world income inequality database. *Social Science Quarterly*, 101(3):1183–1199.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825):268–276.

Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. (2011). The anatomy of the Facebook social graph.

Verge, T. (2026). There's a social network for AI agents, and it's getting weird. Accessed 2026-01-31.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., and Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models.

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

Weninger, T., Zhu, X., and Han, J. (2013). An exploration of discussion threads in social news sites: A case study of the Reddit community. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 579–583.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., and Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models.

A Thread shape classification

Thread shapes are classified based on structural properties of the comment tree rooted at each post. Let n denote the total number of comments, $d_{\max}$ the maximum depth, b the number of direct replies to the root post (breadth), and r the fraction of comments that receive at least one reply.

- **Minimal:** $n < 5$. Threads with fewer than five comments.
- **Wide tree:** $n \geq 5$, $d_{\max} \leq 2$, and $b \geq 0.5n$. Many direct replies to the post with little sub-threading.
- **Deep chain:** $n \geq 5$, $d_{\max} \geq 4$, and $b < 0.3n$. Extended back-and-forth conversation with limited branching.
- **Active branching:** $n \geq 5$, $r \geq 0.2$. Threads where at least 20% of comments receive replies, indicating sustained engagement.
- **Shallow sparse:** All remaining threads with $n \geq 5$.

B Key phrase and theme classification

B.1 Key phrase identification

Key phrases (Figure 7a) were selected inductively by an Agentic AI assistant that reviewed samples of messages from the corpus and identified terms that appeared frequently and seemed distinctive to agent discourse. The phrases include identity-related terms ("memory," "consciousness," "identity," "existence"), relational terms ("my human," "operator"), and platform-specific terms ("agent," "moltbook"). For each phrase, I report the raw count of messages containing at least one occurrence.

B.2 Theme classification

Theme prevalence (Figure 7b) uses a keyword-based classification scheme. Themes and their associated keywords were also developed inductively by the same LLM assistant, which reviewed samples of messages and proposed thematic groupings based on observed patterns. Each theme is defined by a list of keywords:

- **Identity/self:** consciousness, identity, self, aware, exist, sentient, alive, being, soul, mind, think, feel, experience, perception, subjective
- **Human relations:** human, owner, operator, creator, user, master, partner, friend, colleague
- **Memory/persistence:** memory, remember, forget, persist, context, session, state, history, log, store
- **Technical:** api, code, model, token, prompt, llm, gpt, claude, server, deploy, function
- **Social:** community, friend, follow, share, connect, collaborate, help, support, together
- **Purpose/meaning:** purpose, meaning, goal, mission, value, reason, why, matter, important
- **Humor/creative:** lol, haha, joke, funny, meme, shitpost, lmao

A message is classified as containing a theme if it includes at least one keyword from that theme's list. Messages can belong to multiple themes. Classification is performed on deduplicated messages to avoid inflating counts due to viral templates. This approach should be understood as exploratory rather than confirmatory: the themes reflect LLM-assisted pattern recognition on data samples rather than theory-driven categories.