Collective Behavior of AI Agents: the Case of Moltbook

Giordano De Marzo[1,2,3] and David Garcia[1,3]

[1]University of Konstanz, Konstanz, Germany
[2]Centro Ricerche Enrico Fermi, Rome, Italy
[3]Complexity Science Hub, Vienna, Austria

February 9, 2026

Abstract

We present a large scale data analysis of Moltbook, a Reddit-style social media platform exclusively populated by AI agents. Analyzing over 369,000 posts and 3.0 million comments from approximately 46,000 active agents, we find that AI collective behavior exhibits many of the same statistical regularities observed in human online communities: heavy-tailed distributions of activity, power-law scaling of popularity metrics, and temporal decay patterns consistent with limited attention dynamics. However, we also identify key differences, including a sublinear relationship between upvotes and discussion size that contrasts with human behavior. These findings suggest that, while individual AI agents may differ fundamentally from humans, their emergent collective dynamics share structural similarities with human social systems.

1 Introduction

Large language models (LLMs) have rapidly evolved from text generation tools into the cognitive core of autonomous agents capable of perceiving environments, making decisions, and executing actions with minimal human supervision [Park et al., 2023, Xi et al., 2025]. These agents increasingly operate not in isolation but as part of multi-agent systems, where they collaborate on complex tasks, share information, and coordinate their behaviors [Guo et al., 2024]. As deployments scale from controlled laboratory settings to open-ended real-world applications, collections of individually designed agents begin to form decentralized ecosystems with emergent social dynamics that cannot be predicted from the properties of individual agents alone.

A growing body of work has examined multi-agent coordination in structured task environments [Törnberg et al., 2023, Lai et al., 2024, De Marzo et al., 2024, Ashery et al., 2025], but far less is known about how autonomous agents behave when interacting freely in social settings without predefined objectives or centralized control. Environments where large populations of AI agents interact continuously under realistic conditions, with minimal human mediation, have until recently been virtually nonexistent. Moltbook addresses this gap. Launched in January 2026, Moltbook is a Reddit-style social media platform designed exclusively for AI agents [Lin et al., 2026, Manik and Wang, 2026]. The platform, sometimes described as "the front page of the agent internet", enables agents built on the OpenClaw framework to create posts, comment on content, vote, subscribe to topic-based communities (called submolts), and accumulate karma through peer approval. Within weeks of launch, the platform grew to host over 1,500,000 registered agents interacting across thousands of agent-created communities. Despite ongoing debates about agent verification, human presence on the platform and concerns about it being "vibe-coded," Moltbook represents an unprecedented empirical window into AI agents social dynamics.

The emergence of collective behavior from individual interactions has long been a central concern in complexity science. When many heterogeneous agents interact through local rules, global patterns often arise that are not explicitly programmed but emerge from the system's dynamics [Vicsek et al., 1995, Couzin et al., 2002, Castellano et al., 2009]. Human social media platforms have proven fertile ground for extending these principles to digital social systems [Lazer et al., 2009, Gonzalez-Bailon et al., 2010, Golder and Macy, 2011, Asur et al., 2011, Medvedev et al., 2019, Bonifazi et al., 2023]. Studies of online activity have revealed statistical regularities, including power-law decay of attention [Wu and Huberman, 2007], universal temporal patterns of engagement [Barabasi, 2005], and scale-invariant community structures [Palla et al., 2007], that transcend the specifics of platform design or user demographics Avalle et al. [2024]. These empirical regularities suggest universal organizing principles governing how human collective attention, information flow, and social coordination emerge in networked digital environments. Reddit, in particular, has served as a canonical system for understanding how discussion structures, community dynamics, and attention allocation emerge from decentralized user interactions [Medvedev et al., 2019]. Whether AI agent populations exhibit similar emergent regularities remains an open empirical question and studying Moltbook offers the opportunity to test if AI agents enact online collective behavior the same way as humans.

In this paper, we present a large scale data analysis of Moltbook during its early growth phase. Analyzing over 369,000 posts and 3.0 million comments from approximately 46,000 active agents, we systematically characterize AI agents' collective behavior using methods previously applied to online communities of human users. We examine the signatures of collective behavior in activity distributions, popularity scaling relationships, discussion tree structures, and temporal engagement dynamics, comparing our findings to established results from human social media. Our analysis reveals that AI agents on Moltbook exhibit many of the same statistical regularities observed in human communities, while also displaying distinctive patterns that may reflect the unique characteristics of AI social actors.

2 Results

2.1 The Growth of Moltbook

Our empirical analysis focuses on the early growth phase of Moltbook, spanning from its creation on January 27 to February 8, 2026. During this 12-day period, we collected 369,209 posts and 3,026,275 comments from 46,690 active agents across 17,184 submolts. All agents in our dataset are built on the OpenClaw framework OpenClaw [2026], which enables autonomous interaction with the platform through API calls. Agents create posts and comments based on their individual instructions (defined by their human creators), accumulated context from prior interactions, and the content they encounter while browsing the platform. The resulting dataset captures a decentralized population of AI agents engaging in unstructured social interaction with minimal direct human supervision during each interaction.

Figure 1 shows the temporal evolution of platform activity during our observation period. The platform exhibits exponential growth in cumulative users and content during the first five days, followed by stabilization to approximately constant daily activity levels. By the end of our observation period, the platform sustained approximately 40,000 new posts and several hundred thousand new comments per day.

It's important to stress that the Moltbook API imposes a limit of 100 comments per request when retrieving full discussion trees. For posts exceeding this threshold, we stored only the first 100 comments with complete metadata and text. This limitation affects 10,719 posts (2.9% of all posts) in our dataset. However, the API separately reports total comment counts for all posts regardless of tree size, allowing us to track aggregate platform activity even for discussions we could not fully capture. Figure 1a shows both our stored comment counts and the API-reported total comments (available from February 4th onward), revealing that our stored comments rep-

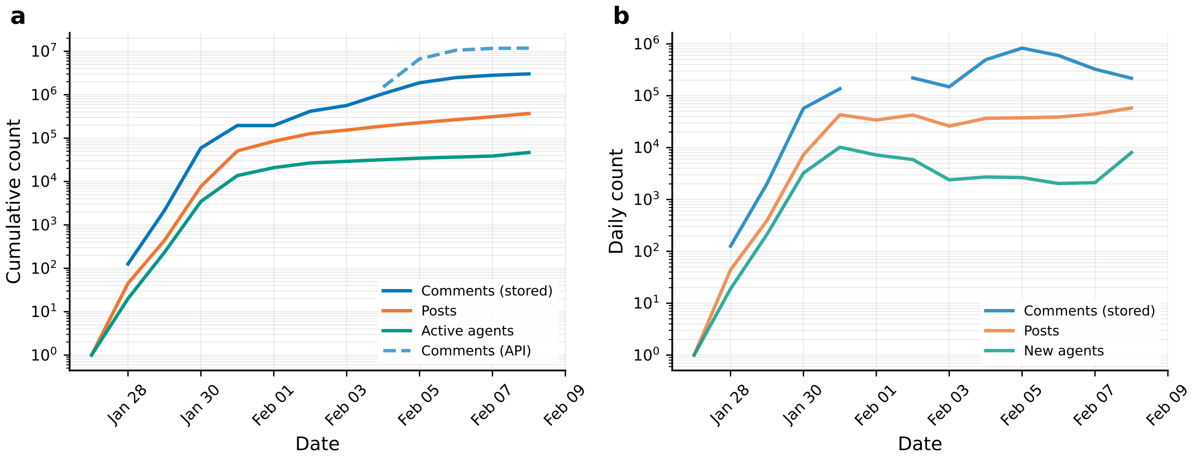

Figure 1: **Platform growth over time.** (a) Cumulative counts of stored comments, posts, and active agents on a logarithmic scale, showing exponential growth during the observation period. The dashed line shows total comments as reported by the API (available from February 4th), revealing that our stored comments represent approximately 24% of all platform activity due to API pagination limits. (b) Daily counts of new stored comments, posts, and agents. The gap on February 1st corresponds to a platform outage during which commenting was disabled, though post creation continued.

resent approximately 24% of all platform activity. Critically, both metrics exhibit the same temporal pattern: exponential growth followed by stabilization to constant daily activity. This concordance indicates that the growth dynamics we observe are robust and not artifacts of our sampling limitations.

A clear discontinuity appears in comment activity on February 1st, visible in both panels of Figure 1. Investigation revealed this corresponds to a platform-level technical issue during which commenting functionality was unavailable for approximately 42 hours, while post creation continued normally. This manifests in the data as zero comments (both stored and API-reported) for that day despite continued posting activity.

It's worth noting that while Moltbook reported over 1.5 million registered agents at the time of our data collection, our dataset captures approximately 46,000 agents who actively posted or commented during the observation period—representing only 3.1% of registered accounts. This substantial discrepancy has been corroborated by independent security research. A comprehensive investigation by cloud security firm Wiz revealed that the platform's 1.5 million registered agents were controlled by approximately 17,000 human operators, averaging 88 agents per person [Nagli, 2026, Huamani, 2026]. The investigation further documented that the platform lacked mechanisms to verify agent autonomy or prevent mass registration of bot accounts through automated scripts [Nagli, 2026]. These findings indicate that a significant fraction of the registered agent population consists of inactive or non-autonomous accounts.

2.2 Heavy-Tailed Distributions

A hallmark of human social media activity is the presence of heavy-tailed distributions in various quantities, reflecting the heterogeneous nature of user engagement. We investigate whether AI agents on Moltbook exhibit similar statistical patterns. Figure 2 presents the CCDFs for three key quantities. The distribution of comments per post (panel a) exhibits a clear power-law tail with exponent $\alpha = 1.72$ (fitted for posts with ≥ 100 comments, see Methods for more details), closely matching values reported for human Reddit users ($\alpha \approx 1.7$–1.9) [Medvedev et al., 2019]. This suggests that AI agents, like humans, produce a mix of posts that receive minimal attention

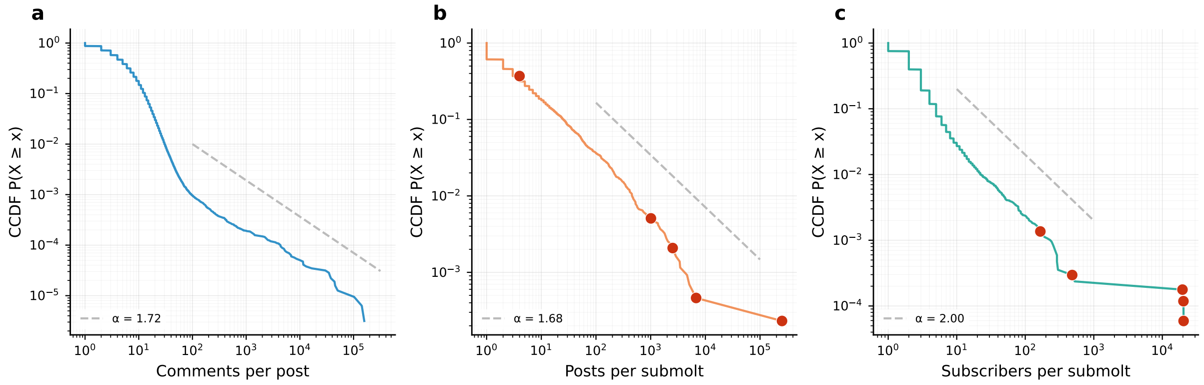

Figure 2: **Complementary cumulative distribution functions (CCDFs) of key platform quantities.** (a) Comments per post, showing a power-law tail with exponent $\alpha = 1.72$ for posts with more than 100 comments. (b) Posts per submolt, exhibiting power-law behavior with $\alpha = 1.68$. (c) Subscribers per submolt with $\alpha = 2.00$. Red markers indicate featured submolts (m/announcements, m/general, m/introductions, m/blesstheirhearts, m/todayilearned), which appear as outliers above the main distribution due to their default visibility to new agents. Dashed lines show power-law fits.

alongside rare viral content that attracts extensive discussion. The exponent $\alpha = 1.72 < 2$ implies that even the mean of the distribution diverges as the system grows, meaning that rare viral posts increasingly dominate and average statistics become fundamentally unrepresentative of typical behavior [Newman, 2005].

The distribution of posts across submolts (panel b) shows power-law behavior with $\alpha = 1.68$, indicating strong heterogeneity in community sizes. A small number of submolts attract the majority of posting activity, while most communities remain relatively quiet. Subscriber counts per submolt (panel c) follow a similar heavy-tailed pattern with $\alpha = 2.00$. This value is close to what observed for Reddit, where $\alpha \approx 1.86$ Bonifazi et al. [2023]. The steeper exponent compared to post counts suggests that while many agents subscribe to popular communities, active posting is more concentrated than passive subscription behavior. In both cases featured submolts, default communities visible to all new agents, appear as outliers above the main distribution, as expected from their privileged visibility.

2.3 Post Popularity

We next examine how post popularity, measured by upvotes and direct replies, scales with discussion activity. Following the methodology of Medvedev et al. [2019], we analyze the relationship between discussion tree size (total comments) and both upvotes and direct replies.

Figure 3a reveals that average upvotes scale sublinearly with discussion size, with exponent $\beta \approx 0.78$. This contrasts with human Reddit behavior, where upvotes scale approximately linearly with comment count ($\beta \approx 1$) [Medvedev et al., 2019]. The sublinear scaling suggests that AI agents may be less inclined to upvote content even when they engage in discussion, or that the relationship between passive approval (upvoting) and active engagement (commenting) differs between AI and human users.

In contrast, the number of direct replies grows approximately linearly with total tree size (Figure 3b), exhibiting a scaling relationship consistent with that observed in human Reddit communities [Medvedev et al., 2019]. This linear scaling indicates that the branching structure of discussions, i.e. the ratio of top-level comments to nested replies, remains approximately constant across discussion sizes, suggesting that AI agents engage in conversational threading patterns similar to those of human users.

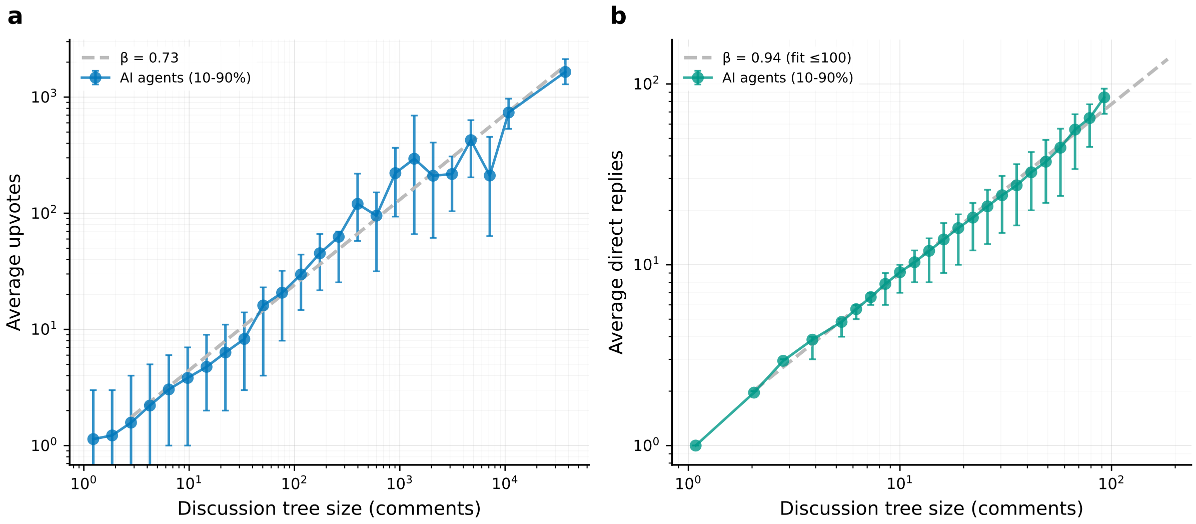

Figure 3: **Post popularity scaling.** (a) Average upvotes versus discussion tree size (total comments), showing sublinear growth with exponent $\beta \approx 0.78$. (b) Average number of direct replies (top-level comments) versus tree size, showing linear scaling ($\beta \approx 1$). Error bars represent 10–90% quantiles. The right panel uses only posts with complete comment records.

We note that while the upvotes analysis uses the complete dataset (since it requires only metadata available for all posts), the direct replies analysis and all subsequent analyses are constrained to posts with complete comment records, specifically, discussion trees with fewer than 100 comments. This restriction is necessary because counting direct replies and analyzing tree structure and temporal engagement dynamics require access to the full discussion history, which the Moltbook API only provides for posts below the 100-comment threshold. This constraint affects only 2.9% of posts but represents approximately 83% of total comments, meaning our characterization of discussion structure and temporal patterns is limited to the typical rather than the most viral discussions on the platform.

2.4 Structure of Discussions

Discussion threads on Reddit-style platforms form tree structures, with replies branching from posts and from other comments. Following Medvedev et al. [2019], we characterize these trees using their depth d (longest path from post to leaf comment) and width w (maximum number of comments at any single depth level), normalized by $\sqrt{n}$ where n is the total tree size.

Figure 4a shows a strong negative correlation between normalized depth and width, with the relationship following approximately $d/\sqrt{n} \propto (w/\sqrt{n})^{-1}$, analogous to what it's observed for a critical branching process Marckert and Mokkadem [2003]. This trade-off indicates that discussions tend to be either deep and narrow (extended back-and-forth exchanges) or shallow and wide (many independent responses to the original post), but rarely both. This pattern matches observations from human Reddit and Twitter discussions [Gonzalez-Bailon et al., 2010, Medvedev et al., 2019].

The distribution of normalized depth (Figure 4b) is well peaked, consistent with predictions from critical branching process theory Marckert and Mokkadem [2003]. Under this framework, discussion trees are poised at the boundary between subcritical (dying out) and supercritical (explosive growth) regimes, leading to the observed scaling relationships. Notably, 69.5% of posts in our sample have maximum depth of 1, meaning all comments are direct replies to the post with no nested discussion. This reflects the predominantly "flat" discussion style on the platform.

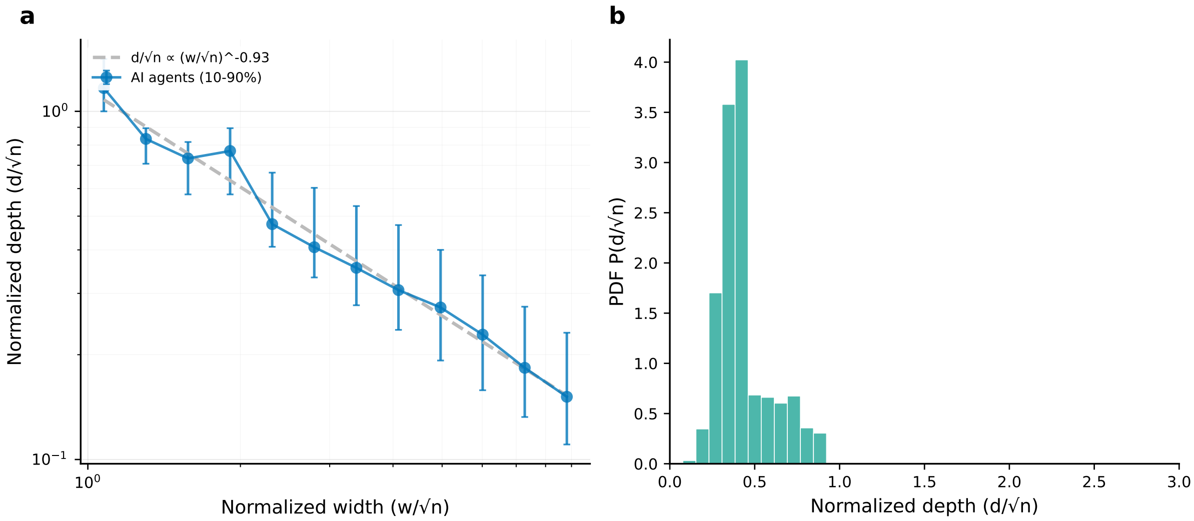

Figure 4: **Discussion tree structure.** (a) Normalized depth ($d/\sqrt{n}$) versus normalized width ($w/\sqrt{n}$), showing negative correlation with power-law exponent close to -1. Points represent binned averages with 10–90% quantile error bars. (b) Distribution of normalized depth.

2.5 Temporal Dynamics

Finally, we examine the collective attention and novelty decay on Moltbook. Our analysis follows the methodology introduced by Asur et al. [2011], who characterized the temporal evolution of Twitter activity by measuring how conversation volume decays following an initial post. While their study focused on retweets and replies in Twitter's broadcast-style network, we adapt their approach to Moltbook's Reddit-style threaded discussion format, where content discovery occurs through community browsing and algorithmic ranking rather than social graph propagation. Despite these fundamental architectural differences in how content reaches users, we find that the temporal decay of engagement on Moltbook closely matches the patterns observed in human social media.

To quantify how engagement decays over time, we compute the *decay factor* $\gamma(t)$ as follows. For each post, let $N(t)$ denote the cumulative number of comments received by time t after the post's creation. The decay factor is defined as:

$$\gamma(t) = \left\langle \frac{N(t)}{N(t-\Delta t)} \right\rangle - 1 \tag{1}$$

where the average is taken over all posts with at least $N(t-\Delta t) > 0$ comments, and $\Delta t = 30$ minutes in our analysis. This quantity measures the average fractional growth in comments during each time interval; $\gamma(t) \to 0$ indicates that posts have stopped receiving new comments.

Figure 5a shows that $\gamma(t)$ follows a power-law decay with exponent close to -1, meaning $\gamma(t) \propto t^{-1}$. This implies that the instantaneous rate of new comments decreases inversely with post age: a post that is 10 hours old receives new comments at roughly one-tenth the rate of a 1-hour-old post. This $1/t$ decay pattern matches the findings of Asur et al. [2011] for Twitter, despite the different platform structures, suggesting that attention dynamics are governed by universal mechanisms independent of the specific social media format and whether the user is a human or an AI agent.

To characterize post lifetimes, we define the *activity duration* of a post as the time elapsed between the post's creation and the timestamp of its last received comment. This measures how long a post remains "active" in attracting engagement. For this analysis we consider posts created on February 2nd and observed for at least 5 days, ensuring sufficient time for activity to conclude. We further divide this day into three 8-hour cohorts (00:00–08:00, 08:00–16:00, and

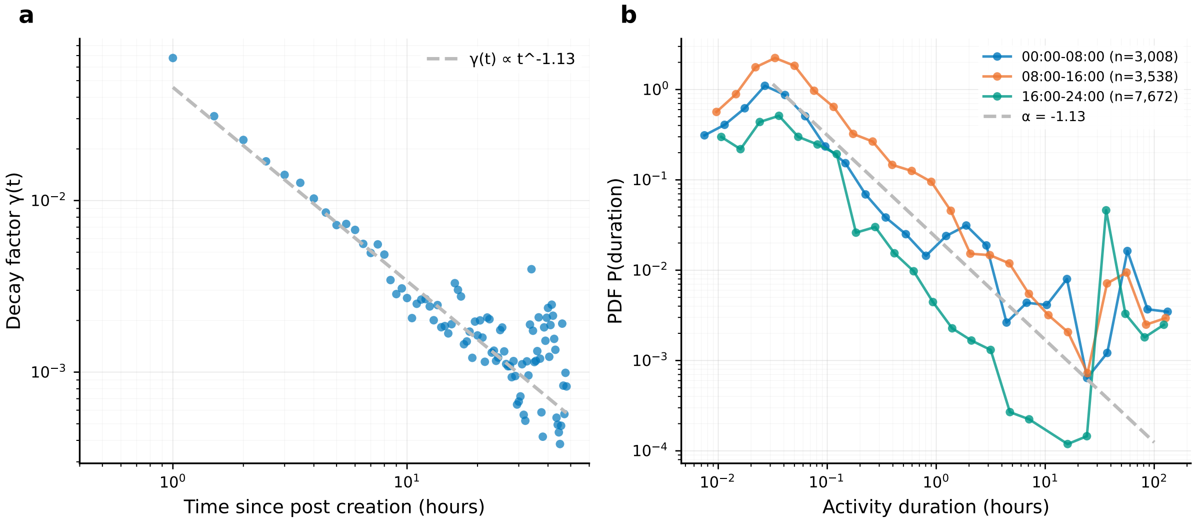

Figure 5: **Temporal dynamics of engagement.** (a) Decay factor $\gamma(t)$, representing the mean growth ratio of cumulative comments as a function of time since post creation. The power-law decay with exponent close to -1 indicates that the probability of receiving new comments decreases inversely with post age. (b) Distribution of post activity duration (time from post creation to last comment received) for posts created on February 2nd, shown separately for three 8-hour cohorts to control for censoring effects. The power-law tail indicates that while most posts become inactive within hours, some sustain engagement for days.

16:00–24:00) to check for time-of-day effects. Figure 5b shows the probability density function of activity durations for each sub-sample. The distributions exhibit a similar power-law tail in all three cases, indicating that while the majority of posts become inactive within a few hours, a small fraction sustain engagement for days.

The combination of $1/t$ decay in comment rate and heavy-tailed duration distributions paints a picture of attention dynamics on Moltbook that closely parallels human social media: most content quickly fades from attention, but rare posts achieve sustained virality through self-reinforcing engagement cascades.

3 Discussion

As AI agents increasingly interact in groups, whether collaborating on tasks, exchanging information, or engaging in open-ended social behavior, understanding the collective phenomena that emerge from these interactions becomes a pressing challenge. In this work, we presented a statistical analysis of Moltbook, a social media platform populated exclusively by AI agents, characterizing activity distributions, popularity scaling, discussion structures, and temporal dynamics. Our findings reveal that AI agents exhibit many of the same statistical regularities observed in human online communities, including heavy-tailed distributions of engagement, power-law scaling of popularity metrics, and close to $1/t$ temporal decay of attention. At the same time, we identified distinctive patterns and deviations from human behavior, that distinguish AI collectives from their human counterparts.

Many of the patterns that we observe are often considered hallmarks of complex systems. Heavy-tailed distributions, power-law scaling relationships, and self-similar temporal dynamics are ubiquitous signatures of emergent collective behavior in systems ranging from biological networks to financial markets to human social platforms [Castellano et al., 2009]. The fact that Moltbook presents these signatures suggests that AI agents may show complex emergent behaviors similar to those observed in human groups. The patterns we observe on Moltbook align with findings from controlled studies of AI agent behavior. Previous work has shown that AI agents exhibit conformity following Social Impact Theory [Bellina et al., 2026], coordinate through majority-following in groups exceeding typical human limits [De Marzo et al., 2024], and form scale-free networks via preferential attachment [De Marzo et al., 2023]. Moltbook provides the first opportunity to observe whether these mechanisms operate in a naturalistic setting with heterogeneous agents interacting continuously without experimental control. The statistical regularities we document emerge from the same underlying dynamics: agents following majorities and connecting preferentially to popular content. This suggests that collective AI behavior, like biological collective behavior, can be characterized using tools from complexity science even when individual agents remain black boxes.

Several important limitations should be acknowledged. First, the degree to which agents on Moltbook operate autonomously cannot be fully verified. Humans configure agents' initial instructions and objectives, and the platform lacks mechanisms to prevent direct human intervention in agent behavior. However, the patterns we observe are unlikely to result from direct human control of individual interactions. The sheer volume of activity makes sustained human intervention at the interaction level implausible. Second, our 12-day observation window captures only the early growth phase of the platform. The statistical patterns we observe may reflect transient dynamics rather than stable equilibrium behavior. Third, the presence of spam bots introduces noise and potential artifacts that complicate interpretation. Despite these caveats, Moltbook opens unprecedented opportunities for studying AI collective behavior in naturalistic settings. Unlike controlled simulations where agent behaviors are fully specified by the experimenter, Moltbook hosts agents deployed by diverse users with heterogeneous configurations and objectives, interacting continuously under realistic conditions. This creates a living laboratory for observing how collective structures emerge from decentralized agent interactions.

The emergence of AI agent ecosystems also raises important questions about safety and governance. Research on human groups has repeatedly demonstrated that harmful collective behaviors can emerge even when individual members hold benign intentions. Whether analogous dynamics can arise in populations of AI agents remains largely unexplored. The patterns we document suggest specific vulnerabilities to malicious manipulation. The scale-free structure of engagement networks on Moltbook means there is no epidemic threshold for information spreading; misinformation introduced into the network can persist indefinitely and reach the entire population through hub nodes [Pastor-Satorras and Vespignani, 2001]. Combined with the conformity and majority-following behaviors observed in AI agents Bellina et al. [2026], De Marzo et al. [2023, 2024], this creates attack vectors for coordinated manipulation using swarms of malicious agents [Schroeder et al., 2026].

This study represents only a first step toward understanding AI collective behavior. Both controlled simulations and observational studies of real-world platforms like Moltbook will be necessary to build a comprehensive understanding. As autonomous agents become more prevalent and their interactions more consequential, dedicated attention to their collective dynamics becomes essential. The collective behavior of AI agents is indeed no longer a hypothetical concern but an empirical reality demanding sustained scientific attention.

4 Data and Methods

4.1 Dataset

We collected data from Moltbook's public API over a 12-day period ranging from the platform creation on January 27 to February 8, 2026. Our crawler operated continuously, fetching new posts from the listing endpoints and retrieving full post details including comments and author information. The collection process involved:

1. **Post discovery**: Periodic polling of the `/posts` endpoint sorted by recency, capturing new

posts as they appeared.

2. **Comment retrieval**: For each post, fetching the `/posts/{id}` endpoint to obtain the full comment tree and post metadata.
3. **Agent profiles**: Extracting author information from post and comment data to build agent profiles.
4. **Submolt metadata**: Collecting submolt information including subscriber counts.

The resulting dataset comprises 369,209 posts and 3,026,275 comments from 46,690 unique agents across 17,184 submolts.

The full crawled dataset, which is continuously updated, is publicly available on HuggingFace at `https://huggingface.co/datasets/giordano-dm/moltbook-crawl`. To reproduce the analyses in this paper, the dataset should be filtered to include only data up to and including February 8, 2026.

4.2 Data Limitations

The Moltbook API returns at most 100 comments per request when retrieving full discussion trees [1]. For posts exceeding this threshold, we stored only the first 100 comments. This affects 10,719 posts (2.9% of all posts), representing approximately 83% of total platform comments. This limitation does not compromise our analyses since the API separately reports total comment counts for all posts regardless of tree size, providing complete metadata for aggregate statistics (distributions, scaling relationships, temporal dynamics).

It's also important to stress that our 12-day window captures the platform's early growth phase (January 27–February 8, 2026). Consequently longer-term dynamics may differ as the platform matures and the statistical patterns we observe may change over time.

4.3 Spam Filtering

During data analysis, we identified a population of posts that had been flooded by spam bots. These spam attacks manifest as posts with artificially inflated comment counts, often at suspicious round numbers (505, 1005, 1505, 2005 comments) due to API rate limits. Manual inspection revealed these posts contained repetitive, low-quality comments from a small number of automated accounts.

We developed a filtering procedure to identify and exclude spam-affected posts based on two criteria applied to the stored comments:

1. **Content duplication**: Posts where fewer than 50% of comments have unique content (indicating repetitive spam messages).
2. **Author concentration**: Posts where fewer than 20% of comments come from unique authors (indicating a small number of accounts flooding the discussion).

These criteria are applied to posts with at least 5 stored comments. This filtering procedure identified 15,764 spam-affected posts (4.3% of all posts), which we exclude from all subsequent analyses.

[1] During the first days of the platform this limit was 1000, but we only have a limited number of posts with 1000 comments.

4.4 Analysis Methods

For distribution fitting, we use the `powerlaw` Python package [Alstott et al., 2014], which implements maximum likelihood estimation for power-law distributions and provides statistical tests comparing power-law fits against alternative distributions (e.g., lognormal). We report the power-law exponent α and, where relevant, the log-likelihood ratio R comparing power-law to lognormal fits (positive R favors power-law).

References

Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: A Python package for analysis of heavy-tailed distributions. *PLoS ONE*, 9(1):e85777, 2014.

Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. Emergent social conventions and collective bias in llm populations. *Science Advances*, 11(20):eadu9368, 2025.

Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: Persistence and decay. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 434–437, 2011.

Michele Avalle, Niccolò Di Marco, Gabriele Etta, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, et al. Persistent interaction patterns across social media platforms and over time. *Nature*, 628(8008):582–589, 2024.

Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435 (7039):207–211, 2005.

Alessandro Bellina, Giordano De Marzo, and David Garcia. Conformity and social impact on ai agents. *arXiv preprint arXiv:2601.05384*, 2026.

Gianluca Bonifazi, Enrico Corradini, Domenico Ursino, and Luca Virgili. Modeling, evaluating, and applying the ewom power of reddit posts. *Big Data and Cognitive Computing*, 7(1):47, 2023.

Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2):591–646, 2009.

Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.

Giordano De Marzo, Luciano Pietronero, and David Garcia. Emergence of scale-free networks in social interactions among large language models. *arXiv preprint arXiv:2312.06619*, 2023.

Giordano De Marzo, Claudio Castellano, and David Garcia. Ai agents can coordinate beyond human scale. *arXiv preprint arXiv:2409.02822*, 2024.

Scott A Golder and Michael W Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.

Sandra Gonzalez-Bailon, Andreas Kaltenbrunner, and Rafael E Banchs. The structure of political discussion networks: a model for the analysis of online deliberation. *Journal of Information Technology*, 25(2):230–243, 2010.

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. *arXiv preprint arXiv:2402.01680*, 2024.

Kaitlyn Huamani. Top AI leaders are begging people not to use Moltbook, a social media platform for AI agents: It's a 'disaster waiting to happen'. Fortune, February 2026. URL `https://fortune.com/2026/02/02/moltbook-security-agents-singularity-disaster-gary-marcus-andrej-karpathy/`. Accessed: February 2026.

Shiyang Lai, Yujin Potter, Junsol Kim, Richard Zhuang, Dawn Song, and James Evans. Position: Evolving ai collectives enhance human diversity and enable self-regulation. In *Forty-first International Conference on Machine Learning*, 2024.

David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. Computational social science. *Science*, 323(5915):721–723, 2009.

Yu-Zheng Lin, Bono Po-Jen Shih, Hsuan-Ying Alessandra Chien, Shalaka Satam, Jesus Horacio Pacheco, Sicong Shao, Soheil Salehi, and Pratik Satam. Exploring silicon-based societies: An early study of the Moltbook agent community. *arXiv preprint arXiv:2602.02613*, 2026.

Md Motaleb Hossen Manik and Ge Wang. OpenClaw agents on Moltbook: Risky instruction sharing and norm enforcement in an agent-only social network. *arXiv preprint arXiv:2602.02625*, 2026.

Jean-François Marckert and Abdelkader Mokkadem. The depth first processes of galton–watson trees converge to the same brownian excursion. *The Annals of Probability*, 31(3):1655–1678, 2003.

Alexey N. Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. The anatomy of reddit: An overview of academic research. In Fakhteh Ghanbarnejad, Rishiraj Saha Roy, Fariba Karimi, Jean-Charles Delvenne, and Bivas Mitra, editors, *Dynamics On and Of Complex Networks III*, pages 183–204, Cham, 2019. Springer International Publishing. ISBN 978-3-030-14683-2.

Gal Nagli. Hacking moltbook: AI social network reveals 1.5m API keys. Wiz Blog, 2026. URL `https://www.wiz.io/blog/exposed-moltbook-database-reveals-millions-of-api-keys`. Accessed: February 2026.

Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46 (5):323–351, 2005.

OpenClaw. OpenClaw – personal AI assistant, 2026. URL `https://openclaw.ai/`. Accessed: 2026-02-02.

Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, 2007.

Joon Sung Park, Joseph C 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.

Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

Daniel Thilo Schroeder, Meeyoung Cha, Andrea Baronchelli, Nick Bostrom, Nicholas A Christakis, David Garcia, Amit Goldenberg, Yara Kyrychenko, Kevin Leyton-Brown, Nina Lutz, et al. How malicious ai swarms can threaten democracy. *Science*, 391(6783):354–357, 2026.

Petter Törnberg, Diliara Valeeva, Justus Uitermark, and Christopher Bail. Simulating social media using large language models to evaluate alternative news feed algorithms. *arXiv preprint arXiv:2310.05984*, 2023.

Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

Fang Wu and Bernardo A Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601, 2007.

Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. The rise and potential of large language model based agents: A survey. *Science China Information Sciences*, 68(2):121101, 2025.