



City Research Online

City, University of London Institutional Repository

Citation: Nudo, J., Cinelli, M., Baronchelli, A. & Quattrociocchi, W. (2025). From Niche to Mainstream: Community Size and Engagement in Social Media Conversations. IEEE Transactions on Computational Social Systems, doi: 10.1109/tcss.2025.3604221

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/36449/>

Link to published version: <https://doi.org/10.1109/tcss.2025.3604221>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

From Niche to Mainstream: Community Size and Engagement in Social Media Conversations

Jacopo Nudo¹, Matteo Cinelli¹, Andrea Baronchelli²,
and Walter Quattrociocchi¹

¹ Center of Data Science and Complexity for Society, Department of
Computer Science, Sapienza University of Rome, Rome, Italy

² Department of Mathematics, City St George's, University of London, and
The Alan Turing Institute, London, UK

Abstract

The architecture of public discourse has been profoundly reshaped by social media platforms, which mediate interactions at an unprecedented scale and complexity. This study analyzes user behavior across six platforms over 33 years, exploring how the size of conversations and communities influences dialogue dynamics. Our findings reveal that smaller platforms foster richer, more sustained interactions, while larger platforms drive broader but shorter participation. Moreover, we observe that the propensity for users to re-engage in a conversation decreases as community size grows, with niche environments as a notable exception, where participation remains robust. These findings show an interdependence between platform architecture, user engagement, and community dynamics, shedding light on how digital ecosystems shape the structure and quality of public discourse.

1 Introduction

Social media platforms have transformed online participation, becoming integral to daily life as primary sources of information, entertainment, and

communication [1–3]. Although these platforms provide unprecedented opportunities for connectivity and information sharing, their integration with complex social dynamics has raised significant concerns about their social impact. A major challenge in studying how platforms influence human behavior is the lack of accessible data [4], and even when researchers obtain data through special agreements with companies such as Meta, it may still be insufficient to distinguish inherent behaviors from the effects of platform design [5–8]. This difficulty arises because the data, deeply embedded in platform interactions, complicate separating intrinsic human behavior from the influences exerted by the platform’s design and algorithms. Research has delved deeply into issues such as polarization, misinformation, and antisocial behaviors in digital spaces [9–13], highlighting the intricate and multifaceted effects of social networks on public discourse.

In this context, a key area of research focuses on identifying differences in dialogue dynamics across platforms, exploring whether these are shaped by algorithmic or design choices, or by community-driven factors [14–16]. The attention a piece of content gets on a platform helps measure its reach, impact, and spread. When users comment repeatedly on a post, it shows they are deeply engaged and willing to share their thoughts [17, 18].

This study compares public conversations, combining a long-term perspective with a multi-platform dataset to examine their dynamics across different digital environments. We define an interaction as the comments a user posts within a thread. To measure user engagement—and, by extension, their attention—we analyze the number of comments left in a thread and the probability of re-entering the conversation after posting, following the methods outlined in [19]. We investigate how these aspects are influenced by the platform where the conversation takes place and by size of the context, defined either as the number of people commenting in a thread or as the number of people that belong to a community or a page where the conversation is hosted.

Looking at the probability re-entry in a thread, that is similar to the probability of leaving more than one comment in a conversation, we notice its sensitiveness to size of the context, analogous to quorum sensing in bacteria, where population density governs collective behaviors through the exchange of signaling molecules [20]; much like how the behavior of individual particles in a system often depends critically on the size of the system itself, as observed in phenomena like finite-size effects in phase transitions and critical phenomena, where collective behavior dominates individual dynamics [21].

So the main research questions are:

- RQ1** Do users maintain consistent participation and re-entry patterns across various platforms?
- RQ2** How does the size of the crowd participating in a conversation affect the probability of user re-entry?
- RQ3** How does the outreach size of the community hosting the conversation impact the probability of user re-entry?

Using data from [14], we systematically compare threads (or conversations) linked to posts on Usenet, Gab, Reddit, Voat, Twitter, and Facebook to answer these questions. The conversations are from a wide range of topics (vaccines, conspiracies, politics, news, Brexit, climate change) and cover 33 years. This research provides actionable insights for platform designers and policymakers seeking to foster healthier online conversations.

2 Related works

The academic literature has explored various aspects of user behavior in online communities.

Studies such as [19] analyzed the likelihood of users re-entering a conversation after leaving a comment, using specific participation patterns as predictors.

Researchers studied the structure of the conversation on Reddit, intended as a tree graph, and analyzed the local and global features, like post sentiment and Subreddit popularity [22]. While [23, 24] investigated how the structure is connected to the toxicity level of the comments produced. Additionally, [18] suggests that persuasiveness increases as comments become more deeply nested in a thread structure.

Works like [25] analyze how users explore diverse options before settling into specific communities, highlighting their persistence within them and their consistency in discussing similar topics despite volatile engagement across different communities. Studies such as [26] and [27] analyze loyalty and turnover among user or consumer in groups or subscription programs, looking for factors that influence it.

In [28], researchers demonstrated how variability in user activity changes significantly with the size of the community on Reddit, suggesting implications for platform design. Relatedly, literature on the attention economy investigated human attention [29–31] as limited resource that platforms and communities compete to capture and retain.

A comprehensive multi-platform analysis of vocabulary usage is provided by [16], exploring linguistic patterns and variations across various social media platforms. Finally, recent studies such as [14] have conducted multi-platform analysis to identify consistent patterns in toxic discourse dynamics, highlighting persistent patterns in toxic behavior across various platforms. Despite the limited material, these studies provide a solid starting point for analyzing attention dynamics, offering a comprehensive overview of user behavior in online communities that can be used as a solid background for further analysis.

3 Materials and methods

This section outlines the data sources for all analyzed social media platforms, detailing the data acquisition process, the preprocessing steps, and the measures employed in the analysis.

3.1 Data Collection

Usenet: Usenet is one of the oldest communication networks on the internet, established in 1980. It is organized into a decentralized system of newsgroups, each dedicated to a specific topic or theme. Users post messages or “articles” to these newsgroups, creating public threads for others to view, read, and reply to. Due to its unmoderated and decentralized nature, Usenet is known for hosting a variety of niche discussions, some of which predate modern social networks. We collected data for the Usenet discussion system by querying the Usenet Archive (<https://archive.org/details/usenet?tab=about>). We selected a list of topics considered adequate to contain a large, broad and heterogeneous number of discussions involving active and populated newsgroups. As a result, we selected conspiracy, politics, news, and talk as topic candidates for our analysis.

Facebook: Facebook is a widely used social networking platform launched in 2004 that enables users to create personal profiles, add friends, and join

groups organized by interests. Facebook centers around individual profiles and direct relationships, facilitating the sharing of posts, images, and comments among friends, as well as reactions such as likes. Users interact within groups or public pages to discuss topics of interest, share information, or engage with communities. We use datasets from previous studies that covered discussions about news [32] gathered from a list of pages by the Europe Media Monitor that reported the news in English.

Twitter: Twitter (now rebranded X) is a social networking service where users post and interact through short messages known as "tweets." Users can follow others, retweet, reply, like, and engage in discussions around hashtags or trending topics. Twitter's timeline focuses on user-generated content organized around individual user accounts, interactions, and hashtags rather than groups or communities. We used a list of datasets from previous studies that includes discussions about vaccines [33], climate change [34] and news [35] topics.

Gab: Gab is a social media platform launched in 2016, marketed as a "free speech" alternative to more mainstream networks. It is organized similarly to Twitter, with users following one another and posting status updates, comments, and reacting to other users' posts. Gab has become popular among users banned or discontent with other platforms due to content moderation policies [36]. We collected data from the Pushshift.io archive (<https://files.pushshift.io/gab/>) concerning public discussions.

Reddit: Reddit is a social content aggregation website, organized in communities constructed around specific topics, named subreddits. Each user has an account corresponding to a user name used to post submissions or to comment on other submissions and other comments. In addition, users can also upvote or downvote a submission in order to show their appreciation or criticism for it. Differently from other social media, Reddit's homepage is organized around subreddits and not on user-to-user relationships. Therefore, subreddits chosen by users are likely to represent their preferred topics and the main source of information consumed on the website. We collected public data from Reddit using the Pushift collection [37].

Voat: Voat.co was a news aggregator website similar in structure to Reddit, operating until December 25, 2020. It gained attention as a migration hub for users banned from Reddit, hosting discussions in specialized communities called "subverses." Users could subscribe to subverses of interest and interact with content through comments, upvotes, and downvotes. Given the platform's alignment with certain controversial topics, Voat's user

community became a focal point for discussions that were censored on other platforms. For this study, we utilized a dataset collected from [38], which captures a snapshot of interactions on Voat prior to its shutdown.

3.2 Preprocessing

According to the social media description provided in the previous section, each platform provides post-related discussion in the form of comments threads. For each platform, such comment threads were extracted and for each comment, the *timestamp* of the publication, the *post_id* (i.e the thread identifier), and the *user_id* were obtained. If the *user_id* was not available due to anonymization, the comment was not included. Table 1 shows a data breakdown.

	Users	Comments	Threads
Gab	166,833	12,509,891	3,764,443
Reddit	394,733	1,798,628	808,016
Twitter	13,620,442	22,337,801	325,451
Usenet	212,259	3,566,773	682,362
Voat	153,255	3,454,791	413,854
Facebook	17,700,372	39,732,512	1,000,000

Platform	Time Range
Gab	Oct 2016 to Oct 2018
Reddit	Jan 2018 to Dec 2022
Twitter	Jan 2010 to Jan 2023
Usenet	Feb 1989 to Jun 2013
Voat	Nov 2013 to Dec 2020
Facebook	Dec 2009 to Aug 2016

Table 1: Data breakdown.

3.3 Measure of Users Participation

In order to study how the number of unique participants varies with conversation length, we provide a measure of users’ tendency to participate in

comments threads. Following the method used in [19], we construct a two-dimensional *density matrix* that describes the distribution of user participation across conversations of different lengths. Given a list of comments, where u is the group of users who produced the comments, let $d = |u|$ represent the number of unique users, and let k denote the number of comments in the list. For each value k , where k ranges from 0 to 200, we select all threads $t_{k,p} \in T_p$, where T_p represents the set of threads on platform p , that contain at least k comments. Again for each value of k , we select among $t_{k,p}$ just the first k comments for each thread, and we calculate the relative frequency of threads having d users where $d \in [1, k]$. These relative frequencies are contained in a density matrix D , where each cell $c_{d,k}$ holds the value $P(d|k)$, the probability that within the first k comments there are d unique users. Note that column-wise the matrix D contains discrete probability and thus each column sums up to one. The patterns that may emerge from the density matrix highlight different participation dynamics. For example, a distribution near the diagonal indicates a “broad-spectrum conversation,” where new users frequently contribute a single comment each. Conversely, density concentrated towards the bottom of the matrix suggests a “concentrated participation,” indicating that a few users produce most of the comments. The second case reflects intense discussions among a small group of interlocutors, that stand for longer attention and reading time.

3.4 Measure of User Re-entry

To assess the activity of a user posting multiple comments and engaging in dialogue, we choose a measure that evaluates the concentration associated with the distribution of the number of users making k comments within a thread. This measure is based on the second and fourth moments of the distribution. Defined k as the interaction length, we compute the following metric:

$$L = \frac{(\sum_k \phi_k^2)^2}{\sum_k \phi_k^4}$$

Where ϕ_k represents the probability that a user writes k comments under the same thread. This measure, L , is intended to quantify the degree of concentration in the distribution of comments, and the position of the peak [39]. A higher value of L indicates that users tend to leave about L

comments on the same thread, while a low value, close to 1, indicates a general habit of leaving just one comment and leaving the conversation. This measure provides insight into user engagement within discussions, helping us to understand how participation varies among users and whether certain conversations present users with different levels of commitment. It also provides a mathematical basis for comparing the distribution of comments across different threads or posts. Similar techniques based on higher moments of distributions have been used in the study of social network behaviors [40].

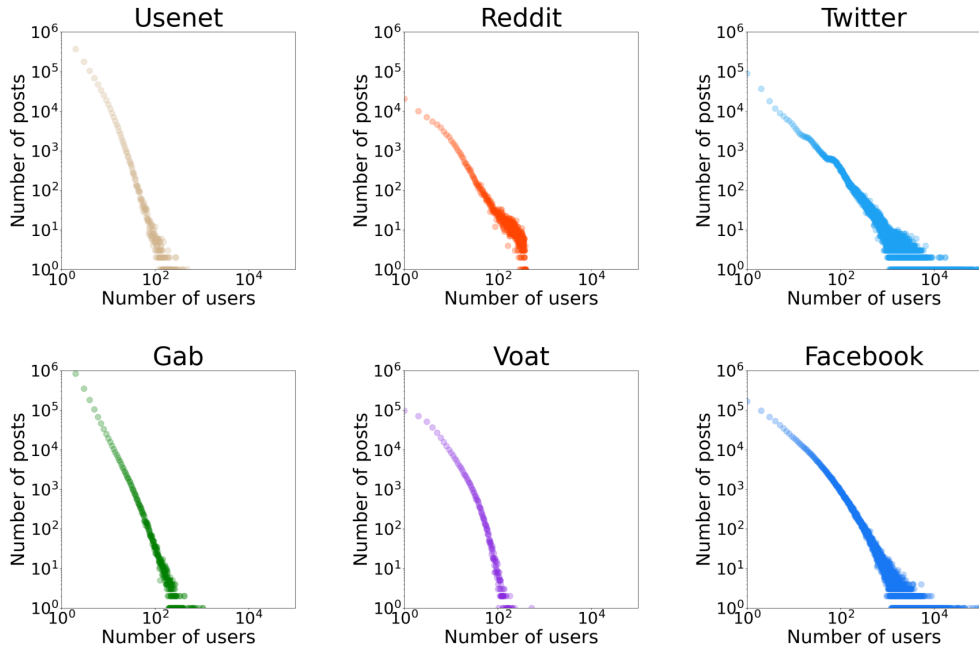


Figure 1: Log-log distribution of the number of unique users commenting under a post, across different platforms. The y-axis represent the absolute frequency of posts that have this amount of unique user. Note the heavy-tailed distributions typical of power-law behaviors.

3.5 Measures of Size

To investigate how individual user engagement, measured by re-entry probability, is influenced by the size of the social context in which they are embed-

ded, we define size using two distinct metrics: (i) the number of users who have contributed to a comment thread, referred to as the *crowd*, and (ii) the number of individuals visiting the page or group hosting the conversation at a given time, termed *outreach*.

The first way to define the size is using the number of unique users who have participated in a comment thread (crowd). Given u the list of all the users that commented in a specific thread, the size of the crowd is calculated as $d = |u|$. So for each interaction, done by a user in a thread, we can determine the size of the crowd in the thread. Then we can sort by d , and divide all interactions in bins, for each bin calculate the localization index (L), as explained in 3.4 on the distribution of the interaction length ϕ .

The second way to define the size is as the number of users involved in the community of a page or a group that host the conversation, defining it as the outreach. Following [31], we start counting the number of users commenting on a certain week. Given p a page (or a group) and t a specific date, we define $O_p(t)$ as the outreach of that page, calculated as the number of users who have commented during that week, smoothed by a rolling average over a time window of 12 weeks. For each interaction done on each page at time t , we can determine the number of active users on that page p at that time using $O_p(t)$. Defining an interaction as the set of comments written by the same user under a thread, we sort all the interactions based on the outreach (O), and divide them into bins of 1000 observations. Finally, for each bin, we calculate the localization index (L), as explained in 3.4, to determine whether the typical number of comments, or the re-entry probability of a user, are connected to the level of outreach of the page commented.

4 Results and Discussions

4.1 Distribution of Number of Users Across Platforms

Figure 1 displays the distribution of the number of unique users per post on the 6 social media platforms: Gab, Reddit, Twitter, Usenet, Voat, and Facebook. The x-axis represents the number of unique users, and the y-axis represents the number of posts being commented by x users, both plotted on logarithmic scales.

Across all platforms, the distributions exhibit a power-law-like behavior (See Supplementary Information), characterized by a heavy-tailed shape.

This suggests that while most posts receive comments from only a small number of users, a small proportion of posts achieve significantly higher engagement. Such distributions are typical in online platforms [14], where content virality and user behavior follow non-uniform patterns [41].

Platforms within the same columns in the figure show similar interaction patterns. Notably, Twitter and Facebook share broader distributions with extended tails that reflect the occurrence of highly viral posts with significant user interactions.

4.2 Analyzing User Participation

To address **RQ1**, we investigate the heterogeneity of participation across a set of comments and the likelihood of re-entering a conversation on various platforms. The density matrices for user participation on Reddit, Twitter, Usenet, Voat, Facebook, and Gab illustrate how conversation length correlates with the number of unique users, represented by the probability distribution $P(d | k)$, as shown in Figure 2. Here, k denotes the prefix length applied to the number of comments in a thread, while d represents the number of unique users contributing, as explained in 3.3.

Facebook and Twitter exhibit "expansionary" dynamics, where the density trends along the diagonal. This pattern suggests that as the prefix k increases, a comparable number of unique users d engage. Such dynamics indicate broad participation and diverse user interaction within conversations. This comparison underscores different engagement styles: broad, transient participation on platforms like Facebook and Twitter versus deeper, concentrated discussions on Reddit, Voat, Gab, and Usenet.

Thread-wise the re-entry patterns, measured using the number of comments by the same user under a post, is interpretable by applying the localization L to the distribution of number of comments per user (calculated for splitting by week the datasets), as explained in 3.4. From Figure 3, we note a certain degree of platform dependency in Localization distributions. Consistent with previous observations, Facebook exhibits the narrowest distribution ($L \simeq 1$), followed by Twitter. This observation is further supported by a Kruskal-Wallis test ($H = 1539.84$, $p < 0.001$), which indicates significant differences across platforms. A post-hoc Dunn test with Bonferroni correction revealed that most pairwise comparisons were significant, with exceptions observed between Gab and Voat ($p = 0.173$) and Gab and Twitter ($p = 0.311$).

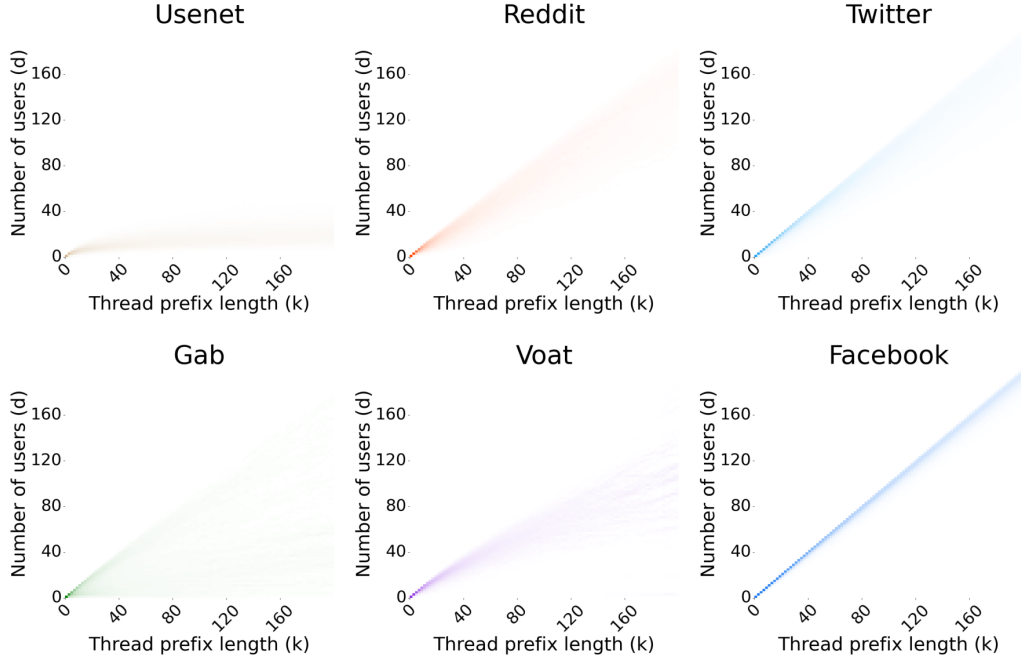


Figure 2: Density matrices illustrating the distribution of a number of unique users (d), given thread prefix length (k , number of comments), across different platforms. The diagonal trend on platforms like Facebook and Twitter indicates a proportional increase in unique user engagement with thread length, while the lower concentration on platforms suggests a smaller core group driving most of the activity. Color intensity reflects the probability densities (conditioned by column), highlighting platform-specific interaction patterns.

Lower values of L can be interpreted as a reduced likelihood of leaving more than one comment under the same post. This measure appears to be distributed similarly across the entire analyzed time span; however, it seems to depend on the number of users participating in the conversation.

In order to obtain further evidence with respect to that provided by analyzing the index L , we calculated the median of the probability distribution that a user-post interaction consists of a single comment for each platform. The medians are as follows: for Usenet, 0.68; for Gab, 0.83; for Reddit, 0.78; for Voat, 0.82; for Twitter, 0.87; and for Facebook, 0.94. (See Supplementary Information).

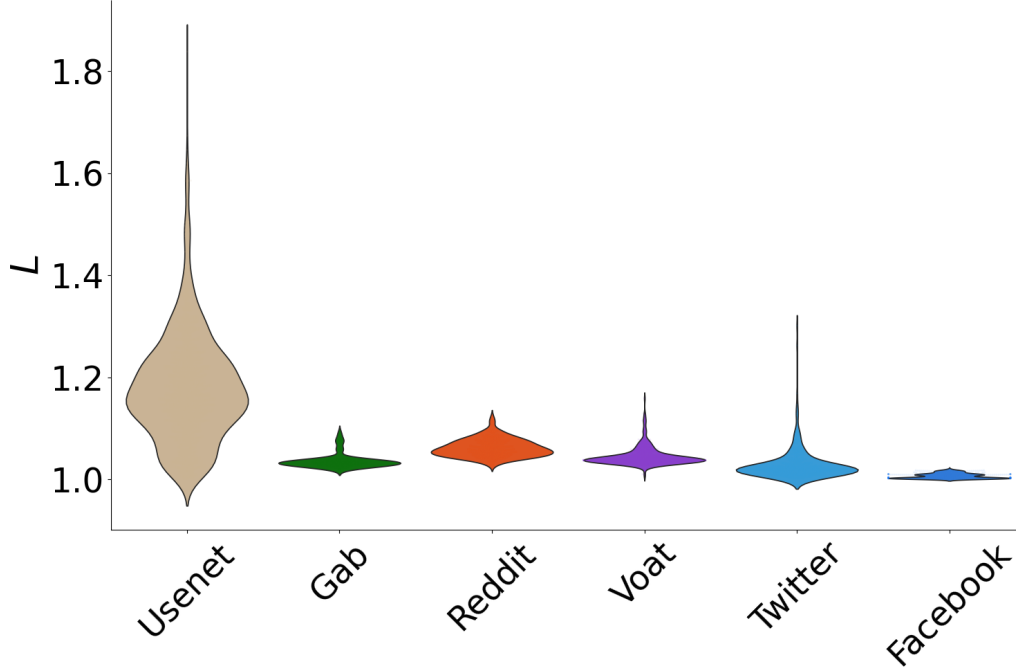


Figure 3: Distribution of the localization parameter (L) across different platforms. The violin plot shows, using L , probabilities that an interaction (among a user and a post) is composed by a certain amount of comments, with each platform exhibiting distinct patterns. Localization values greater than 1 indicate a greater propensity to leave more than one comment in a conversation. While a value close to 1 stand for a distribution with less degrees of freedom.

4.3 User Re-Entry vs Crowd Size

In order to answer **RQ2**, we define size of the crowd commenting under a thread, as explained in 3.5, in order to study its impact on re-entry probability. Once divided all the interactions, i.e. defined an interaction as the set of comments that a user writes under a thread, into bins with respect to the number of users involved (crowd size) we plotted the distribution of the localization (L) parameter for each bin, Figure 4. For Gab and Usenet, it is possible to observe how, as the number of users in a conversation increases, the typical number of comments shifts away from localization values equal to one. While on Reddit and Voat, user engagement initially increases

with the number of participants, peaking around 150 users. Beyond this threshold, additional participants reduce marginal engagement, likely due to conversational noise or decreased cohesion.

Our findings reveal a bound in platforms like Reddit and Voat, where the likelihood of re-entering a conversation peaks around 150 participants. Surprisingly, this quantity aligns with the Dunbar Number, which represents the cognitive limit of stable social connections [42, 43]. This boundary observed also in other domains such as news consumption [44], app-usage [45], and points of interests [46], suggests that conversational dynamics in these environments are bounded not only by platform design but also by cognitive and social factors intrinsic to human behavior. On smaller platforms like Usenet and Gab, we observe a linear increase in user engagement as the number of participants grows. Conversely, larger platforms such as Facebook and Twitter exhibit independence from community size, likely due to thread structures that promote shallow interactions. These differences may stem from platform design choices that either constrain or encourage deeper dialogues. Smaller, niche platforms provide a focused environment, while mainstream platforms prioritize scalability, often at the expense of sustained user engagement.

4.4 User Re-Entry vs Outreach Size

In order to answer **RQ3**, we investigate whether user propensity for dialogue is related to the size of the community associated to a page or a group (outreach size); we follow the methods described in 3.5. On each platform, we analyze user-post interactions and categorize each interaction in a bin of 1000 observations based on their outreach (O) level. To assess the impact of the average page outreach on bin’s localization parameter (L), we perform a linear regression analysis. This approach allows us to examine how the distribution of re-entry varies as a function of increasing page outreach.

As shown in Figure 5, there appears to be a general trend across most platforms where the propensity re-entry in a conversation (i.e., to post multiple comments) is negatively correlated with the outreach size of the community that host the thread. This is due to an increasingly noisy and chaotic environment, where it becomes difficult to follow the thread of the conversation, leading to a more passive and distracted interaction.

This trend is confirmed by all platforms except Twitter and Usenet. On Twitter user behavior appears to be largely unaffected by the size of the

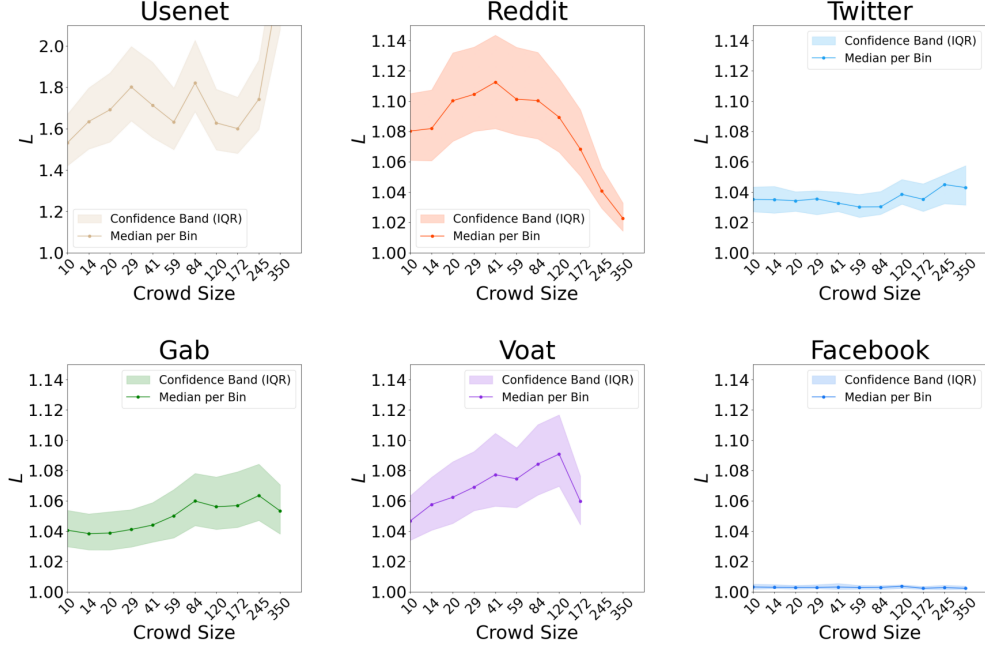


Figure 4: Distribution of the localization parameter (L) across different platforms, controlling by the number of users involved in the thread (crowd size on x-axis). The number of users involved seems to influence the distribution of the number of comments posted by each user under the same post. Linearly for Usenet and Gab, with a saturation effect on Reddit and Voat, while it does not seem to impact giant mainstream platform such as Twitter and Facebook.

outreach, with a general tendency to avoid extended dialogue. In the second case, data suggest that individuals are more likely to persist in conversations as the number of users increases and thus the stimulus.

5 Conclusions

This study analyzes public conversations across six social media platforms over 33 years, uncovering significant differences in conversational dynamics and user engagement. The findings reveal how platform design, community

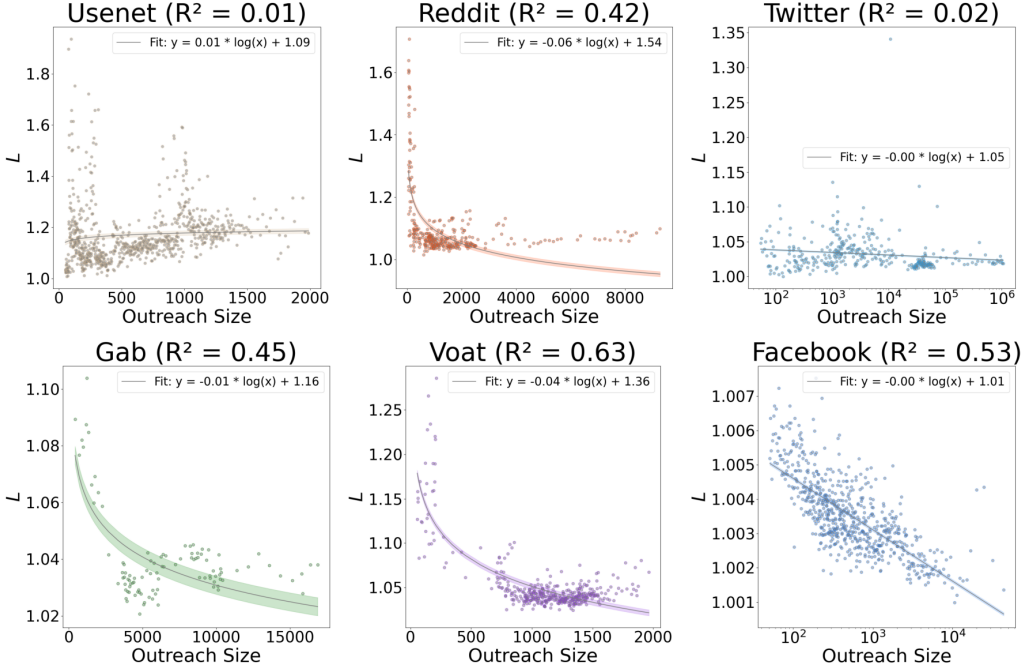


Figure 5: Plot of the localization parameter L (probability of re-entry on y-axis) changing the outreach level of the community (outreach size on x-axis), across different platforms. Values on x-axis are represented in a Log scale for Facebook and Twitter. Each interaction has been grouped in a bin, based on the level of outreach that the community (or the page) hosting the interaction had at the time. It show how, almost everywhere the probability of re-entry in a conversation decrease with the increase of the outreach size, likely due to a progressively noisier conversation environment.

size, and user behavior interact to shape dialogue in digital spaces, providing a nuanced understanding of the factors driving engagement. Responding to RQ1, we find that the propensity for dialogue, measured as the number of re-entries in a conversation, is generally lower on larger, mainstream platforms. These environments favor broader but less sustained participation than smaller, niche platforms. Addressing RQ2, our analysis shows that the propensity for dialogue correlates with the crowd size—the number of users involved in a conversation on smaller platforms. Initially, an increase in crowd size stimulates richer and more diverse interactions. However, conver-

sations fragment and lose cohesion beyond a critical threshold, particularly on platforms like Reddit and Voat. This threshold aligns with cognitive limits, such as Dunbar’s Number, highlighting the role of human constraints on social connections in shaping interaction quality within larger groups. About RQ3, as platform environments scale, user engagement—measured by the likelihood of re-entering a conversation—declines. Across all platforms, larger communities and higher outreach correlate with reduced dialogue persistence. This trend is most pronounced on mainstream platforms like Facebook and Twitter, where content saturation and algorithmic amplification likely drive passive consumption and diminish individual involvement. In contrast, niche platforms maintain higher levels of engagement even as their community size increases, likely due to their focused and less algorithmically driven environments. These findings underscore important implications for platform design and policy-making. By prioritizing scalability and virality, mainstream platforms may inadvertently undermine opportunities for meaningful dialogue, fostering fragmented interactions and passive use. Future research explores how platform policies, user migration, and algorithm changes influence conversational dynamics. Longitudinal studies investigate the link between conversational fragmentation and the effectiveness of policies fostering constructive discourse. In conclusion, public conversations on social media reflect a complex interplay between platform design, community size, and user behavior. Balancing scalability with meaningful interactions is key to fostering constructive dialogue and healthier online environments. Understanding these dynamics helps design better digital spaces for engaging discussions.

References

- [1] Joshua A Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)*, 2018.
- [2] Matthew Gentzkow and Jesse M Shapiro. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839, 2011.

- [3] Thomas Aichner, Matthias Grünfelder, Oswin Maurer, and Deni Jegeni. Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, behavior, and social networking*, 24(4):215–222, 2021.
- [4] Jon Roozenbeek and Fabiana Zollo. Democratize social-media research-with access and funding. *Nature*, 612(7940):404, 2022.
- [5] Sandra González-Bailón, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Deen Freelon, Matthew Gentzkow, Andrew M Guess, et al. Asymmetric ideological segregation in exposure to political news on facebook. *Science*, 381(6656):392–398, 2023.
- [6] Andrew M Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656):398–404, 2023.
- [7] Andrew M Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656):404–408, 2023.
- [8] Brendan Nyhan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. Like-minded sources on facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144, 2023.
- [9] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the national academy of Sciences*, 113(3):554–559, 2016.
- [10] Christopher A Bail, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing

views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018.

- [11] Yonatan Lupu, Richard Sear, Nicolas Velásquez, Rhys Leahy, Nicholas Johnson Restrepo, Beth Goldberg, and Neil F Johnson. Offline events and online hate. *PLoS one*, 18(1):e0278511, 2023.
- [12] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
- [13] Matteo Cinelli, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo. Dynamics of online hate and misinformation. *Scientific reports*, 11(1):22083, 2021.
- [14] Michele Avalle, Niccolò Di Marco, Gabriele Etta, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, and Walter Quattrociocchi. Persistent interaction patterns across social media platforms and over time. *Nature*, 628:582–589, 2024.
- [15] Isaac Waller and Ashton Anderson. Quantifying social organization and political polarization in online platforms. *Nature*, 600(7888):264–268, 2021.
- [16] N Di Marco, Edoardo Loru, Anita Bonetti, Alessandra Olga Grazia Serra, Matteo Cinelli, and Walter Quattrociocchi. Patterns of linguistic simplification on social media platforms over time. *Proceedings of the National Academy of Sciences*, 121(50):e2412105121, 2024.
- [17] Ceren Budak, Brendan Nyhan, David M Rothschild, Emily Thorson, and Duncan J Watts. Misunderstanding the harms of online misinformation. *Nature*, 630(8015):45–53, 2024.
- [18] Lu Xiao and Humphrey Mensah. How does the thread level of a comment affect its perceived persuasiveness? a reddit study. In *Science and Information Conference*, pages 800–813. Springer, 2022.

- [19] Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 13–22, 2013.
- [20] Melissa B Miller and Bonnie L Bassler. Quorum sensing in bacteria. *Annual Reviews in Microbiology*, 55:165–199, 2001.
- [21] H Eugene Stanley. *Phase transitions and critical phenomena*, volume 7. Clarendon Press, Oxford, 1971.
- [22] Yulin Yu, Julie Jiang, and Paramveer S Dhillon. Characterizing the structure of online conversations across reddit. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2):1–23, 2024.
- [23] Martin Saveski, Brandon Roy, and Deb Roy. The structure of toxic conversations on twitter. In *Proceedings of the web conference 2021*, pages 1086–1097, 2021.
- [24] Gabriele Etta, Matteo Cinelli, Niccolo Di Marco, Michele Avalle, Alessandro Panconesi, and Walter Quattrociocchi. A topology-based approach for predicting toxic outcomes on twitter and youtube. *IEEE Transactions on Network Science and Engineering*, 2024.
- [25] Niccolò Di Marco, Matteo Cinelli, Shayan Alipour, and Walter Quattrociocchi. Users volatility on reddit and voat. *IEEE Transactions on Computational Social Systems*, 2024.
- [26] William Hamilton, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. Loyalty in online communities. In *Proceedings of the International AAAI conference on web and social media*, volume 11, pages 540–543, 2017.
- [27] Scott D Mainwaring and Charlotte P Lee. Turnover and the model of coordinated action (moca). In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 239–242, 2017.
- [28] Elliot Panek, Connor Hollenbach, Jinjie Yang, and Tyler Rhodes. The effects of group size and time on the formation of online communities:

- Evidence from reddit. *Social Media+ Society*, 4(4):2056305118815908, 2018.
- [29] Thomas H Davenport and John C Beck. The attention economy. *Ubiquity*, 2001(May):1–es, 2001.
 - [30] Josef Falkinger. Limited attention as a scarce resource in information-rich economies. *The Economic Journal*, 118(532):1596–1620, 2008.
 - [31] Emanuele Sangiorgio, Matteo Cinelli, Roy Cerqueti, and Walter Quattrociocchi. Followers do not dictate the virality of news outlets on social media. *PNAS nexus*, 3(7):pgae257, 2024.
 - [32] Ana Lucía Schmidt, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. Anatomy of news consumption on facebook. *Proceedings of the National Academy of Sciences*, 114(12):3035–3039, 2017.
 - [33] Carlo M Valensise, Matteo Cinelli, Matthieu Nadini, Alessandro Galeazzi, Antonio Peruzzi, Gabriele Etta, Fabiana Zollo, Andrea Baronchelli, and Walter Quattrociocchi. Lack of evidence for correlation between covid-19 infodemic and vaccine acceptance. *arXiv preprint arXiv:2107.07946*, 2021.
 - [34] Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, et al. Growing polarization around climate change on social media. *Nature Climate Change*, 12(12):1114–1121, 2022.
 - [35] Alessandro Quattrociocchi, Gabriele Etta, Michele Avalle, Matteo Cinelli, and Walter Quattrociocchi. Reliability of news and toxicity in twitter conversations. In *International Conference on Social Informatics*, pages 245–256. Springer, 2022.
 - [36] Savvas Zannettou, Barry Bradlyn, Emiliano De Cristofaro, Haewoon Kwak, Michael Sirivianos, Gianluca Stringini, and Jeremy Blackburn. What is gab: A bastion of free speech or an alt-right echo chamber. In *Companion Proceedings of the The Web Conference 2018*, pages 1007–1014, 2018.

- [37] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839, 2020.
- [38] Amin Mekacher and Antonis Papasavva. ” i can’t keep it up.” a dataset from the defunct voat. co news aggregator. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1302–1311, 2022.
- [39] Sidney R Nagel, Gary S Grest, and A Rahman. Phonon localization and anharmonicity in model glasses. *Physical review letters*, 53(4):368, 1984.
- [40] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. Opinion dynamics on interacting networks: media competition and social influence. *Scientific reports*, 4(1):4938, 2014.
- [41] Albert-László Barabási. The physics of the web. *Physics World*, 14(7):33, 2001.
- [42] Robin IM Dunbar. The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 6(5):178–190, 1998.
- [43] Bruno Goncalves, Nicola Perra, and Alessandro Vespignani. Validation of dunbar’s number in twitter conversations. *arXiv preprint arXiv:1105.5170*, 2011.
- [44] Matteo Cinelli, Emanuele Brugnoli, Ana Lucia Schmidt, Fabiana Zollo, Walter Quattrociocchi, and Antonio Scala. Selective exposure shapes the facebook news diet. *PloS one*, 15(3):e0229129, 2020.
- [45] Marco De Nadai, Angelo Cardoso, Antonio Lima, Bruno Lepri, and Nuria Oliver. Strategies and limitations in app usage and human mobility. *Scientific reports*, 9(1):10935, 2019.
- [46] Laura Alessandretti, Piotr Sapiezynski, Vedran Sekara, Sune Lehmann, and Andrea Baronchelli. Evidence for a conserved quantity in human mobility. *Nature human behaviour*, 2(7):485–491, 2018.