

RESEARCH ARTICLE

Bluesky: Network topology, polarization, and algorithmic curation

Dorian Quelle^{1,2*}, Alexandre Bovet^{1,2}

1 Department of Mathematical Modeling and Machine Learning, University of Zurich, Switzerland, **2** Digital Society Initiative, University of Zurich, Switzerland

* dorian.quelle@uzh.ch



Abstract

Bluesky is a nascent “Twitter-like” and decentralized social media network with novel features and unprecedented data access. This paper provides a characterization of its interaction network, studying the political leaning, polarization, network structure, and algorithmic curation mechanisms of five million users. The dataset spans from the website’s first release in February of 2023 to May of 2024. We investigate the replies, likes, reposts, and follows layers of the Bluesky network. We find that all networks are characterized by heavy-tailed distributions, high clustering, and short connection paths, similar to other larger social networks. BlueSky introduced feeds—algorithmic content recommenders created for and by users. We analyze all feeds and find that while a large number of custom feeds have been created, users’ uptake of them appears to be limited. We analyze the hyperlinks shared by BlueSky’s users and find no evidence of polarization in terms of the political leaning of the news sources they share. They share predominantly left-center news sources and little to no links associated with questionable news sources. In contrast to the homogeneous political ideology, we find significant issues-based divergence by studying opinions related to the Israel-Palestine conflict. Two clear homophilic clusters emerge: Pro-Palestinian voices outnumber pro-Israeli users, and the proportion has increased. We conclude by claiming that Bluesky—for all its novel features—is very similar in its network structure to existing and larger social media sites and provides unprecedented research opportunities for social scientists, network scientists, and political scientists alike.

OPEN ACCESS

Citation: Quelle D, Bovet A (2025) Bluesky: Network topology, polarization, and algorithmic curation. PLoS ONE 20(2): e0318034. <https://doi.org/10.1371/journal.pone.0318034>

Editor: Francesco Pierri, Politecnico di Milano, ITALY

Received: August 9, 2024

Accepted: December 29, 2024

Published: February 26, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0318034>

Copyright: © 2025 Quelle, Bovet. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The raw data cannot be shared publicly as per the UZH PhF Ethics Committee’s requirements, as it contains potentially identifying information about users’ social connections, posting patterns, and behavioral data. However, we have deposited

Introduction

Bluesky is a novel and decentralized social media site that opened up in an invite-only beta release in February 2023. The network is a microblogging site, explicitly describing itself as “a Twitter-style social app” [1]. In 2019, Bluesky originated as the “Bluesky initiative” and was announced by the then CEO of Twitter, Jack Dorsey [2]. As a separate entity, the takeover of Twitter (now X) by Elon Musk did not affect Bluesky’s operation. The website has grown to 5.7 million users [3]. In early February of 2024, Bluesky opened the website to users without an invite.

the codes and user IDs that can be used to download all the data necessary to reproduce our results directly from the Bluesky API on an open data repository accessible at <https://doi.org/10.7910/DVN/NGQKDS>. This repository contains a text file of all 4,754,059 valid user DIDs present in the analysis, along with all analysis code including repository download scripts. Using these materials, researchers can reconstruct the complete dataset through standard API requests while complying with Bluesky's Terms of Service.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

While Bluesky is modeled after Twitter, it sets out to solve the “thorniest problems of social media” such as misinformation, harassment, and hate speech by implementing decentralization and leveraging a marketplace approach to these problems[1]. Notably, Bluesky allows anybody to create custom moderation and algorithmic curation services, and users are free to subscribe to the ones they prefer. Decentralization means that the protocol or platform draws upon “multiple interoperable providers for every part of the system”. In practice, it means that several competing clients for the platform such as Graysky [4] and deck.blue [5], and the official Bluesky app are available to each user. Additionally, users can self-host their data on Personal Data Servers (PDS), which store user data and allow other participants of Bluesky to query their data [1]. Bluesky acknowledges that most users will sign up on a shared PDS run by a professional hosting provider. This provider, however, does not need to be Bluesky; it can be run by anyone. Bluesky's decentralized design allows broad data access and the range of choices given to users is unprecedented for a large social media site [6].

In this study, we examine the complete Bluesky network. We investigate users' activity on Bluesky over time and provide the first insights into what drove user sign-ups during the website's rise. Complete and longitudinal studies of a social media platform's evolution are rare as they require extensive time-series data. Consequently, there have been relatively few studies that capture the full developmental trajectory of a social media network from its inception. For example, previous studies have looked at the right-wing social media site Parler and the decentralized platform Mastodon [7] [8]. Here, we characterize the topology of the network of Bluesky throughout the observation period. We describe the network based on various degree distributions, its clustering, density, and connectivity.

In a backlash against “opaque content recommendation systems” [1], multiple decentralized social networks, such as Mastodon, implemented non-algorithmic, reverse chronological feeds [7]. Bluesky sets out to give users more agency over their own user experience. In practice, users have more choices in moderation and can design and subscribe to diverse content recommendation algorithms. Bluesky enables users to generate feeds letting them “choose their algorithms” in an effort to aid users discover content from other users they do not know and to gain exposure to specific content posted. Bluesky's innovative feed feature has been used to create over 39 thousand feeds by a subset of highly active users. Feeds showcases a wide array of algorithmic choices, ranging from simple regex-based filters to professionally curated content streams. Popular feeds include “Discover” which promises to show “Trending content from your personal network” [9], “Mutuals” showing posts from followers of the user, but also topic-specific collections such as “Science” and “Art”. Notwithstanding the breadth of creative feeds, the feature's overall adoption appears limited relative to Bluesky's total user base. Engagement with feeds follows a familiar pattern in social media: a heavy-tailed distribution where a small number of feeds and highly active users dominate the landscape. This aspect of the platform provides an opportunity to investigate how users engage with and potentially influence content curation mechanisms, a topic of growing importance in the study of social media dynamics.

Over the last decade, social media has become more fragmented with an increasing number of smaller or fringe platforms, serving a cohesive group of users [10,11]. Recent years have seen the proliferation of smaller, niche social media platforms catering to specific user groups or ideologies. These platforms often exhibit high levels of homogeneity in terms of user demographics and political leanings, contrary to platforms such as Twitter and Facebook that show a polarization across political ideologies[12,13]. For instance, Gab and Parler have been found to attract predominantly conservative users [8,14–16]. While this homogeneity can create strong communities, it also raises concerns about echo chambers and the

potential for increased polarization. However, an open question is whether, within these more homogeneous platforms, specific issues or topics may still show a polarization of opinions. This study investigates whether Bluesky, as a new and growing platform, exhibits similar patterns of homogeneity observed in other small platforms [8,14–16] or if it manages to attract a more diverse user base. We examine both broad political leanings and specific issue-based discussions to provide an understanding of the platform's user composition and potential for diverse discourse. We look at the spread of misinformation, finding that Bluesky users disseminate little to no information associated with news sources associated with conspiracy theories, propaganda, or fake news. Of all posts containing domains posted to Bluesky only 0.14% are classified as questionable. This small number of posts was authored by 3,704 unique users, making the spread of news marked as questionable almost nonexistent on Bluesky. Next, we investigate the political ideology on a left-right spectrum and show that Bluesky is almost homogeneously left-center biased, with very few users having a right-of-center ideology. Lastly, we show that conversations surrounding the Israel-Palestine conflict are highly polarized, indicating that political homogeneity does not necessarily dictate consensus on specific issues.

Results

Activity on Bluesky

Fig 1 presents the daily number of active users according to six engagement metrics. Dates with the highest numbers of new users over the year 2023 were substantially driven by activity and news about X (formerly Twitter). While further research leveraging qualitative surveys is necessary to establish the exact reasons for users switching, the number of sign-ups significantly correlates with news about Twitter [17].

We look at days with a proportionally high number of new sign-ups by calculating the ratio of new to existing users on the platform over the course of 2023. On September 19, the day with the highest ratio of new to old users, X announced that all users might be charged a fee to use the website ("Elon Musk says Twitter, now X, could charge all users subscription fees" [18], "Elon Musk: Social media platform X, formerly Twitter, could go behind paywall" [19]). September 19 and 20 saw the first and fourth-highest numbers of new active users to existing users on the platform, respectively. The day with the second most sign-ups, relative to the size of the active userbase, was July 3, 2023. On this day, X experienced global outages as a bug caused users to receive rate-limit errors, preventing them from viewing an unlimited number of posts ("Twitter rate-limits itself into a weekend of chaos" [20], "Twitter's Troubles Are Perfectly Timed for Meta" [21]). October 18 and 19 experienced the third and fifth-highest ratios of new users to existing users engaging with Bluesky. On October 18, Twitter announced a \$1 fee for new users in New Zealand and the Philippines ("X, formerly Twitter, rolls out US \$1 annual fee for new users in New Zealand and the Philippines" [22], "Starting today, we're testing a new program (Not A Bot) in New Zealand and the Philippines. New, unverified accounts will be required to sign up for a \$1 annual subscription to be able to post & interact with other posts." [23]). Lastly, on December 21, Twitter experienced another global outage, leading to another surge of sign-ups on Bluesky ("X, formerly Twitter, sees massive outage as tens of thousands report issues" [24], "Is X/Twitter down? Users report problems accessing feeds in multiple countries" [25]). In 2024, the week following Bluesky's opening to the public on the 7th of February 2024 surpassed previous records of new active users compared to the existing user-base. We see very similar patterns across all usage metrics. Following the inception of Bluesky activity slowly grew up to a peak in mid 2023 when activity on the platform slowly decreased. The public opening of Bluesky led to an unprecedented peak with activity

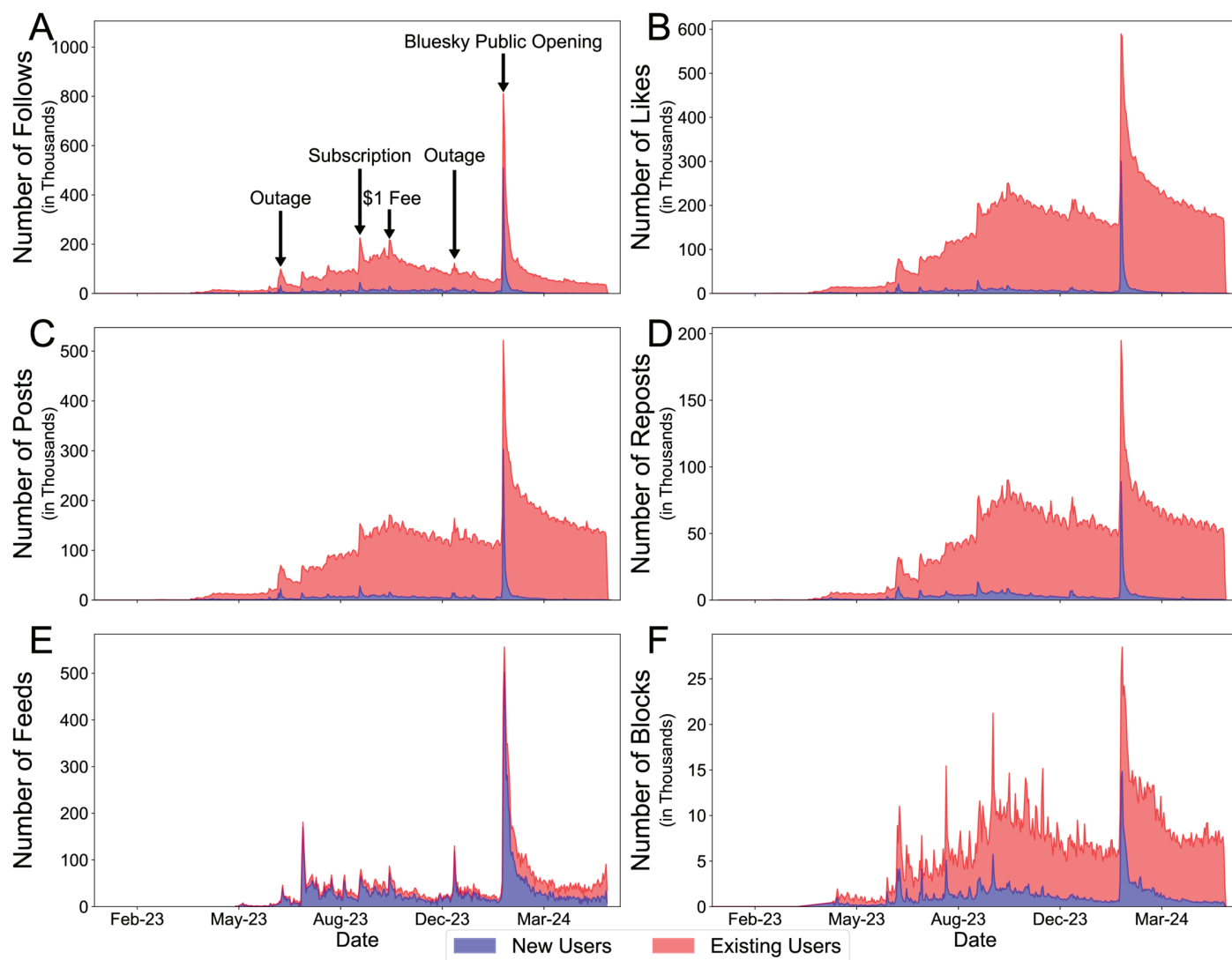


Fig 1. User activity on the Bluesky social media platform from February 2023 to May 20, 2024. Each panel details the number of new and existing active users, ranging from follows (A), likes (B), posts (C), reposts (D), feed generation (E), to blocks (F), showing the number of unique users engaging through these actions. The term 'New Users' refers to individuals interacting for the first time with the platform through the respective activity measure. Blue areas denote new users, while red areas show the number of existing users engaging in activity.

<https://doi.org/10.1371/journal.pone.0318034.g001>

quickly decaying to levels of activity similar to the summer of 2023. Activity on Bluesky now seems to have stabilised for now.

Structure and evolution of the Bluesky network

The data gathered via the Bluesky API represents a temporal network of the entire interaction graph of the social media platform. This allows us to analyze changes in the network topology over time.

Social media sites such as Bluesky are often described as a singular “network” connecting users to each other. However, users of social media sites form relationships and interact with users through a variety of different mechanisms - all capturing different relationships that are

not necessarily ontologically equivalent [26]. Magnani and Rossi [27] find large differences in the centrality of users on social media depending on the interaction layer they investigate. We, therefore, describe the topological structure of Bluesky based on four distinct layers: Followership, Replies, Reposts, and Likes. For a description of the interactions underlying the individual layers, please refer to the Materials and Methods section.

Figs 2 and 3 show the distribution of engagement metrics per post and engagement metrics per user. All distributions exhibit a power-law distribution with a large number of users and posts receiving or authoring few interactions, and a small number of entities being responsible for the vast majority of interactions.

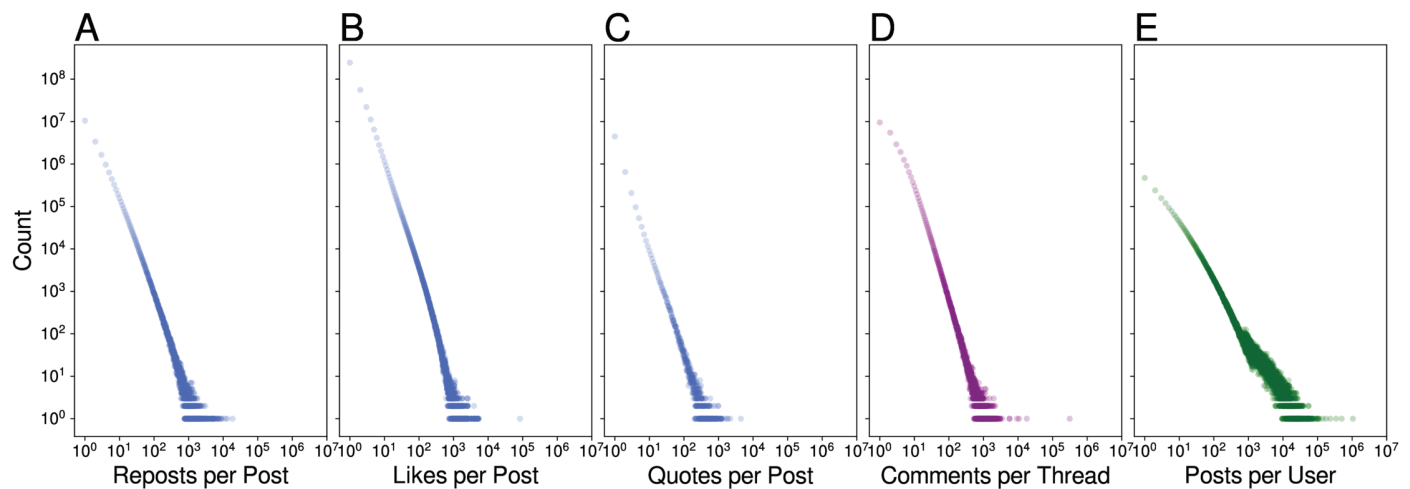


Fig 2. Social media interaction metrics with posts (log-log scale). Each panel plots a specific metric against its frequency to analyze patterns of user engagement and content spread. The X-axis represents the specific metric, and the Y-axis shows the frequency of occurrences for each metric value. (A) Reposts per Post. (B) Likes per Post. (C) Quotes per Post. (D) Comments per Thread. (E) Posts per User. All plots use logarithmic scales.

<https://doi.org/10.1371/journal.pone.0318034.g002>

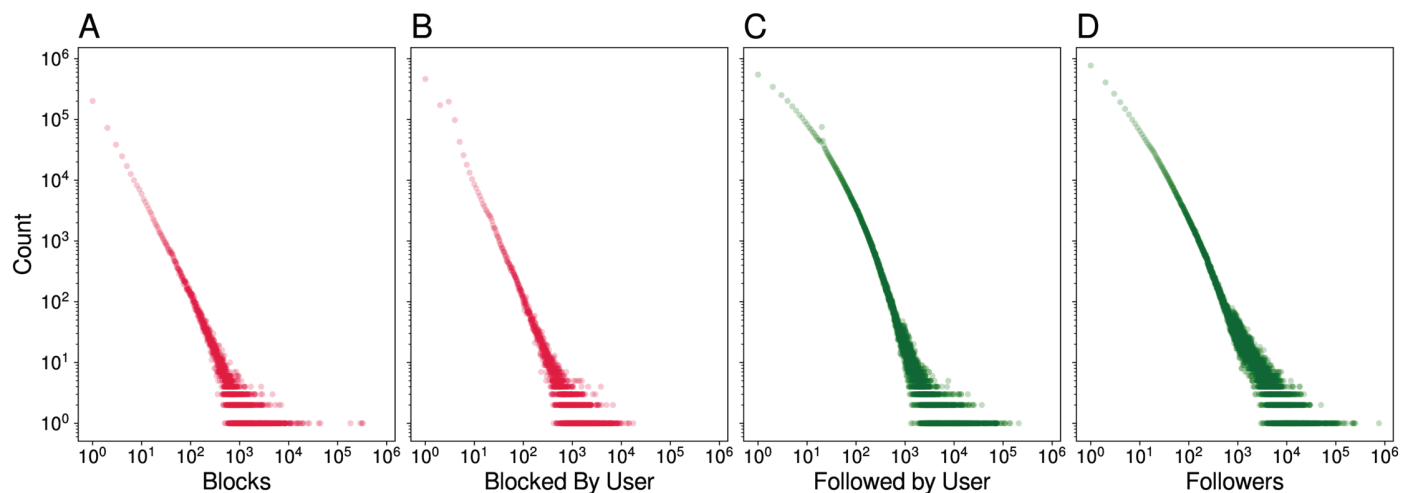


Fig 3. Social media interaction metrics by Users (log-log scale). Each panel represents a specific interaction metric plotted against its occurrence frequency to analyze patterns of user interactions. The X-axis denotes the metric in question, while the Y-axis shows the frequency of occurrences for each metric value. (A) Reposts per Post. (B) Likes per Post. (C) Comments per Thread. (D) Posts per User.

<https://doi.org/10.1371/journal.pone.0318034.g003>

For all distributions we report the mean μ , standard deviation σ , skewness γ , kurtosis β , minimum m , maximum M , the exponent of the power-law distribution α and the ratio of mean to maximum $\frac{\mu}{M}$ in Table 1.

Fig 4 shows three key metrics that chart changes in the network structure from 2023 to May 2024, computed across four distinct networks. Figs 4A to 4D illustrate the weekly count of unique, active users for each network. A clear growth trend in user engagement is observed from February 2023 until September 2023, with peak activity observed at different magnitudes across networks—700,000 users in followership, 300,000 in both replying and reposting, and 600,000 in the Likes network. After these peaks, there is a notable decline in activity, which reverses in February 2024 following Bluesky's public launch, allowing unrestricted user access. This results in record-high weekly activities across the networks: 2 million users in followership, 450,000 in replying, 550,000 in reposting, and 1.3 million in Likes. Since the opening of the platform, the number of follows, comments, reposts and likes on the platform has been slowly decreasing to levels last seen before the opening of Bluesky. The steepest drop in activity is seen in the Follows network. This is likely driven by the initial actions of new users who follow suggested profiles without further significant engagement.

Figs 4E to 4H depict the weekly number of interactions within each network, showing trends similar to those of user engagement. The growth in interactions peaks in September 2023 across all networks—6 million in followership, 3.2 million in replying, 2.2 million in reposting, and 21 million in Likes—before decreasing and then surging to new highs in February 2024. The followership interactions show the highest variability, mirroring the bursty nature of sign-ups on the platform.

Lastly, Figs 4I to 4L focus on the average interactions per unique user within each network. The metrics climb until mid-2023, reaching their zenith in April for Followership with 26 interactions, May for replying with 21 interactions, April for Reposting with 12 interactions, and July for Likes with 60 interactions. A gradual decrease follows until 2024. Upon Bluesky's public opening, there's a noticeable dip in average interactions for Likes, Reposts, and replying, indicating lower activity levels among newer users compared to the earlier, invitation-only cohort. Conversely, the average interactions in the Followership network

Table 1. User activity distributions are indicative of heavy-tailed behavior. Included metrics are the mean μ , standard deviation σ , skewness γ , kurtosis β , minimum m , maximum M , and the ratio of mean to maximum $\frac{\mu}{M}$ from the distributions depicted in Fig 2 and 3. The exponent of the power law distribution is denoted by α . The notably high values of γ and β indicate a pronounced right-skewed, heavy-tailed nature across all distributions. Furthermore, the exceptionally low $\frac{\mu}{M}$ values further confirm the extensive tail behavior characteristic of these distributions. “#” should be read as Number of.

Distribution	μ	σ	γ	β	m	M	$\frac{\mu}{M} \times 10^3$	α
# Quotes per Post	1.682	6.003	149.259	67646.663	1	4455	0.377	5.441
# Likes per Post	2.148	9.521	5355.320	53468722.557	1	102123	0.021	3.702
# Reposts per Post	4.140	20.348	141.972	70043.802	1	18765	0.221	2.503
Thread Length	4.185	62.836	4780.687	23736718.688	1	312092	0.013	2.145
# People Blocked By	8.155	81.228	70.535	9086.269	1	18234	0.447	2.122
# People Blocked	19.494	838.448	340.712	123106.559	1	324147	0.060	1.900
# People Followed by User	38.952	343.076	203.031	71572.734	1	208288	0.187	1.441
# Followers	43.231	668.036	516.367	498518.321	1	755584	0.057	1.527
# Posts per User	89.601	1061.988	501.547	450864.562	1	1059206	0.085	1.458

<https://doi.org/10.1371/journal.pone.0318034.t001>

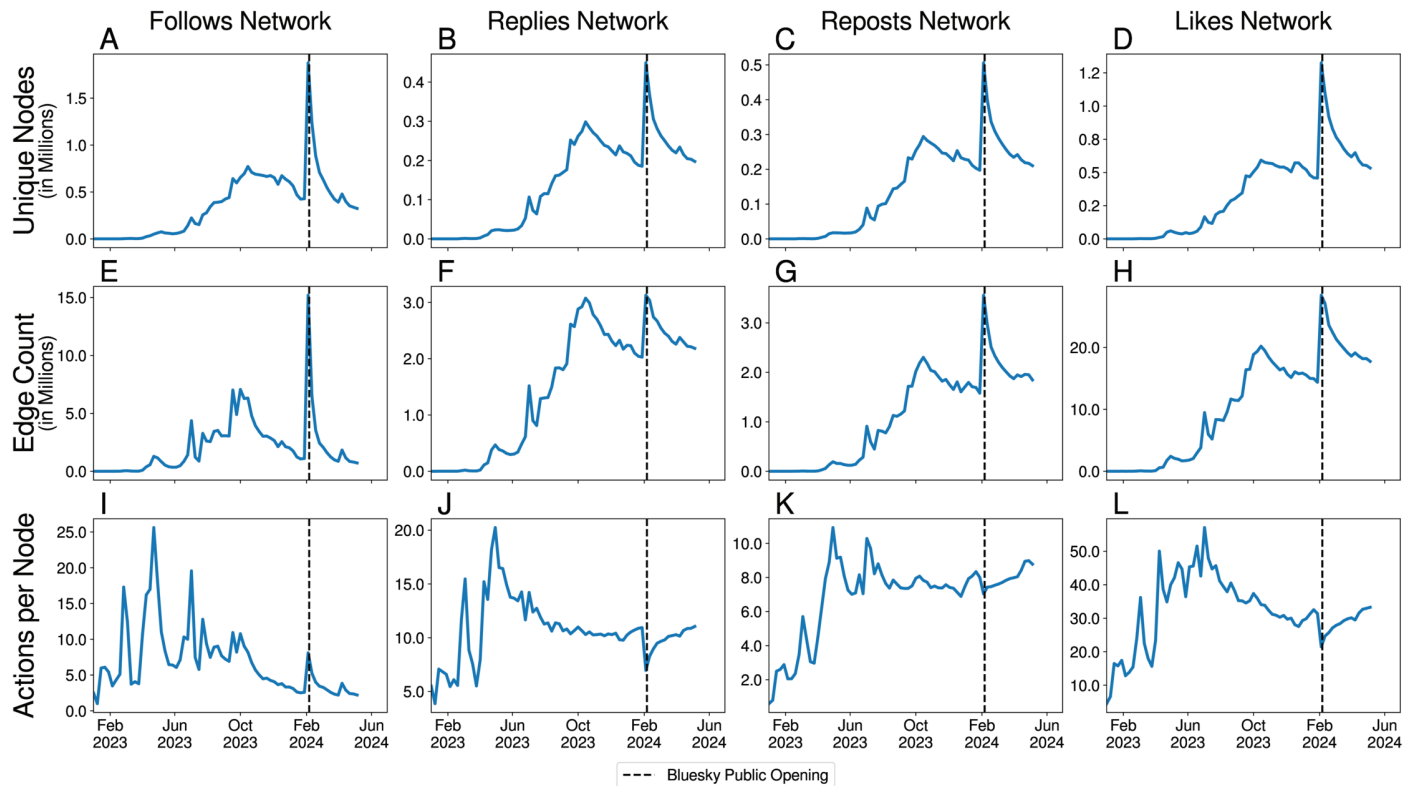


Fig 4. Metrics capturing changes in the network structure from 2023 to May 2024. These metrics are computed across four networks. (A–D) Count of unique nodes (in millions) active per week for each network. (E–H) Number of unique Edges in the network per Week. (I–L) Ratio of edges to unique edges, capturing the activity of nodes in each week. The black dashed line in each graph denotes the date of the public opening of Bluesky.

<https://doi.org/10.1371/journal.pone.0318034.g004>

increase, suggesting that the newly joined users are relatively more engaged in following activities than in other forms of interaction. After the influx of new users the average activity per active user has steadily increased for all but the following network.

Fig 5 shows three measures capturing changes in the structure of the network over the observation period. Fig 5A to Fig 5D focus on the normalized average clustering coefficient for each network, a measure that is adjusted by comparing it to a randomized graph with the same degree sequence. This comparison is visualized where the dashed red line indicates parity between the real and randomized networks. The consistent observation that the normalized clustering coefficient remains above one suggests that the network structure is more cliquish than random models would predict. Similarly to Fig 5A to Fig 5H, the normalized clustering coefficient increases gradually until September of 2023, where activity on Bluesky was locally maximal. The magnitude of peaks varies with the Followership network reaching a coefficient of 10, while Reposts and Likes peak at coefficients of 10 and 16, respectively. The replies network, however, achieves a significantly higher peak of 200, indicating exceptionally dense clustering. Following these peaks, all coefficients trend downwards until February 2024 due to new user influx. While the clustering coefficient remains volatile for the non-persistent interaction, it seems to have somewhat stabilized.

Fig 5E to Fig 5H present the density of these networks over time, which reflects the proportion of actual connections relative to the maximum possible. All density values are plotted on a logarithmic scale to highlight trends more clearly. Despite the non-persistent nature of the

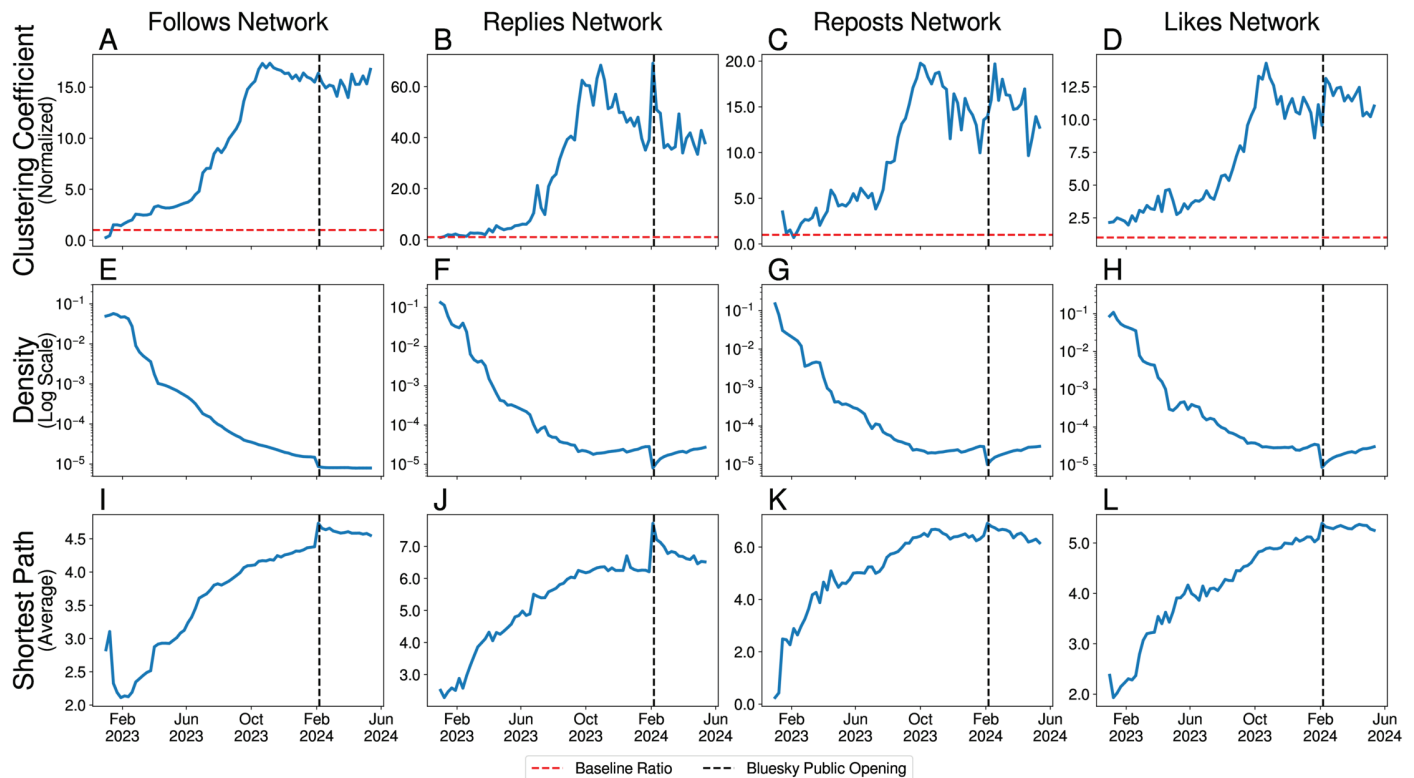


Fig 5. Metrics capturing changes in the network structure from 2023 to May 2024. Clustering-coefficient, density, and average shortest path are computed for four networks. Replies, Reposts, and Likes capture non-persistent interactions, thus all metrics are calculated individually for each week's edges. The followership network is persistent. (A–D) Normalized average clustering coefficient. The dashed red line represents an equal value for the random and original graph. (E–H) Density of the networks. (I–L) Average shortest path length for all networks. The black dashed line in each graph denotes the date of the public opening of Bluesky.

<https://doi.org/10.1371/journal.pone.0318034.g005>

three networks, all exhibit a consistent, sub-linear decline in density over time. This downward trend is accentuated in February 2024, when network density sharply decreases across all networks due to the sudden increase in the user base following Bluesky's public opening.

Fig 5I to 5L show the average shortest path in all networks over time. For all networks we observe a slow and sub-linear increase in the average distance until February 2024. When Bluesky opened up to the public, the average distance sharply increases, to varying extents, across all networks. This can be attributed to new users which are only loosely connected to the network. After these users connected to the network, we observe a slow decline in the average distance for the replies and reposts network, and stagnation in the followership and likes network. Importantly, for all networks the average shortest path remains very low showing the connectivity, efficiency, and small-worldness of the Bluesky network.

Feeds

Unlike traditional social media platforms, Bluesky introduces a novel “marketplace of algorithms” [1] through its feed feature, enabling users to design, implement, and distribute their own content curation mechanisms, ranging from simple keyword matching to complex machine learning models [28]. User-generated feeds are a core functionality of Bluesky and an alternative to “opaque content recommender systems” used by larger platforms. Kleppmann

Table 2. Top Feeds by number of likes on Bluesky.

Displayname	Description	Number of Likes
For You	Learns what you like	16,132
OnlyPosts	Posts from people you follow without ...	5,137
Science	The Science Feed. A curated feed from ...	4,972
Adult Content	Formerly “Suggestive” All (nonviolent) ...	4,180
Art	Images posted by artists on Bluesky...	3,268
New	Posts by furries across Bluesky...	3,221
Mutuals	Posts from users who are following you...	3,103
Home+	Its the Home feed Blue Sky was missing...	3,050
Art	Posts by furries with #furryart...	2,938
Blacksky	Amplifying the voices of any and all...	2,778

<https://doi.org/10.1371/journal.pone.0318034.t002>

et al. [1] cite feeds created by users based on regex matching and machine learning algorithms. Other feeds leverage the network structure of Bluesky and surface content from users' followers. The default algorithm for Bluesky users is a non-algorithmic reverse-chronological feed of their connections.

In total, 39,639 feeds have been created by 18,352 active users showcasing the breadth of content curation algorithms available to users on Bluesky and the broad usage of this novel feature. Users can bookmark a feed [6], which pins the feed on their home screen. While bookmarks are private on Bluesky, it is public knowledge whether users a “liked” a feed. In our dataset, 139,033 Users have used this feature and liked feeds 295,902 times.

The most liked feed *For You* has been renamed *Discover* and promises to show “Trending content from your personal network” [9]. Other popular feeds include “Science”, which is a feed curated by “Bluesky professional scientists” [29]. Other feeds such as the “Hospitality & Tourism 🍍” or “Paleo Sky” use regex patterns to match posts (*tourism*, *skift* and *Paleontology*|*Archaeology*|*#PaleoSky* respectively).

To systematically analyze both the creation and popularity of different content curation approaches, we translate all feed descriptions to English with Google translate and create a topic model of the descriptions. We employed BERTopic [30] for neural topic modeling on the feed descriptions after translating them to English using Google Translate. Feed descriptions were preprocessed by removing URLs, HTML tags, numeric tokens, and non-Latin characters. Using the all-MiniLM-L6-v2 sentence transformer model [31], our analysis identified 463 distinct topics. Of these, 56.84% of feeds were grouped into topics containing multiple feeds, while the remainder were single-feed topics. The largest topics based on feed count and total likes received are detailed in Table 3.

The most prevalent topic category (Topic 0) consists of art-focused feeds ($n = 612$). These feeds aggregate content from artists, and aim to connect artists online. For example, one feed described itself as “Find your artist friends here!”. Similarly, the second most prevalent cluster (Topic 1) consists of users sharing “Music, Songs, and Audio” ($n = 394$). The third largest cluster filters for content on games, particularly board games ($n = 390$). The next topic is related to Japanese pop culture, anime, gaming, and fandom-related feeds, with a particular focus on specific characters, series, or creators, often written in a mix of Japanese romanization and English. Other interesting feed-topics include NSFW content (Topic 4 $n = 354$) and Manga (Topic 9, $n = 330$).

Interestingly, the largest topics by feed count do not correlate strongly with the topics that received the most likes. Only topic 2 ($n_{likes} = 4,260$) is represented in this list. The two largest cluster of feeds based on the number of likes is a cluster entirely composed of “furry” related

Table 3. Top topics by number of feeds created and total likes received. Topics are represented by their most characteristic terms, with counts showing the number of feeds per topic and likes indicating total user engagement with feeds in that topic.

Largest Topic by Number of Feeds			Largest Topic by Number of Likes		
Topic	Representation	Count	Topic	Representation	Likes
0	art, my, artwork, artists, all	612	17	furry, fursuit, furrylist, furries	9,690
1	music, songs, audio, song, jazz	394	346	furry, bskyprobablyawebiste, across	6,222
2	games, game, gaming, board	390	331	follower, follow, followers, back	5,308
3	oshikapu, shobamyu, kawaii, jay	389	2	games, game, gaming, board	4,260
4	nsfw, sfw, content, warning, art	354	21	education, science, teaching	3,881
5	tracking, malifaux, pom, falcom	338	29	books, book, reading, read, readers	3,514
6	tracking, matsuura, nakayama	338	28	likes, sorted, count, liked, hours	3,487
7	words, containing, name, search	337	52	bluesky, posted, compilation, bsky	3,449
8	japanese, fgo, language, japan	333	55	lgbtq, queer, lgbtqia, gay, vibrant	3,327
9	manga, anime, illustrations, drawn	330	58	blue, sky, sorthn, fishing	3,262

<https://doi.org/10.1371/journal.pone.0318034.t003>

feeds. Topic 331 with 5,308 likes consist of feeds which filter content based on social connections and user interactions, like “Posts liked by your follows” or “The last post from each user you follow”. They help users discover content through their social graph and highlight engagement patterns, such as “Posts from your quieter followers” and “Most popular posts by people you follow”. Many of these feeds aim to surface community content and relationship dynamics, like “Posts that are popular with people you follow”. This topic and topic 28 both deal with non-content related filtering. While topic 331 focuses on the social graph of Bluesky, topic 28 shows feeds primarily focused on engagement-based filtering, particularly around likes. These feeds include various ways to surface content based on engagement metrics, from simple filters like “Posts with more than 1000 likes” to more complex sorting like “Most likes and reposts within 24 hours.” Some feeds focus on high engagement (“over 10000 Likes”), while others combine engagement metrics with time windows and language filters. The feeds help users discover trending and highly engaged content through different combinations of these parameters.

Our analysis reveals that users leverage Bluesky’s algorithmic curation capabilities in diverse and creative ways. While traditional engagement-based feeds exist (surfacing highly-liked or trending content), many feeds serve specific community purposes. Some create topic-focused spaces around art, anime, or gaming interests, while others build dedicated safe spaces for marginalized communities (evidenced by popular LGBTQ+ feeds with 3,327 likes). The data shows significant engagement with feeds focused on community building and content discovery, whether through social graph analysis (“Posts liked by your follows”) or interest-based curation (“Find your artist friends here!”). The broad range of moderation approaches on Bluesky enables users to craft these specialized spaces while maintaining platform-wide standards, demonstrating how democratized content curation can effectively serve both broad discovery needs and niche community interests. Our analysis suggests this approach has enabled both broad content discovery and specialized community engagement.

Fig 6 shows the distribution of the number of likes per feed, the number of feeds created per user, and the number of feeds liked per user. Table 2 shows the most liked feeds in our dataset. We again see heavy-tailed distributions with the most active participants liking and creating exponentially more content than the median user.

Table 4 summarizes the descriptive statistics of the distributions. On average, feeds only attracted two likes - showing that most feeds receive little to no engagement. Conversely, users who actively liked feeds liked, on average, over fourteen distinct feeds. We also see that a large number of feeds was created by a small minority of highly active users. This indicates that users who take advantage of this feature express their preferences for algorithmic choices broadly, which could help researchers study algorithmic choices. However, in total, only 139 thousand out of 5 million users liked at least one feed.

Political leaning and polarization of Bluesky

Small and novel social media platforms have oftentimes been characterized by little diversity in political and ideological viewpoints. Truth Social - launched in February of 2022, leans strongly conservative - and was created as an “alternative social media platform” targeting Republican social media users [14]. Voat.co, a small Reddit-esque social network, grew after

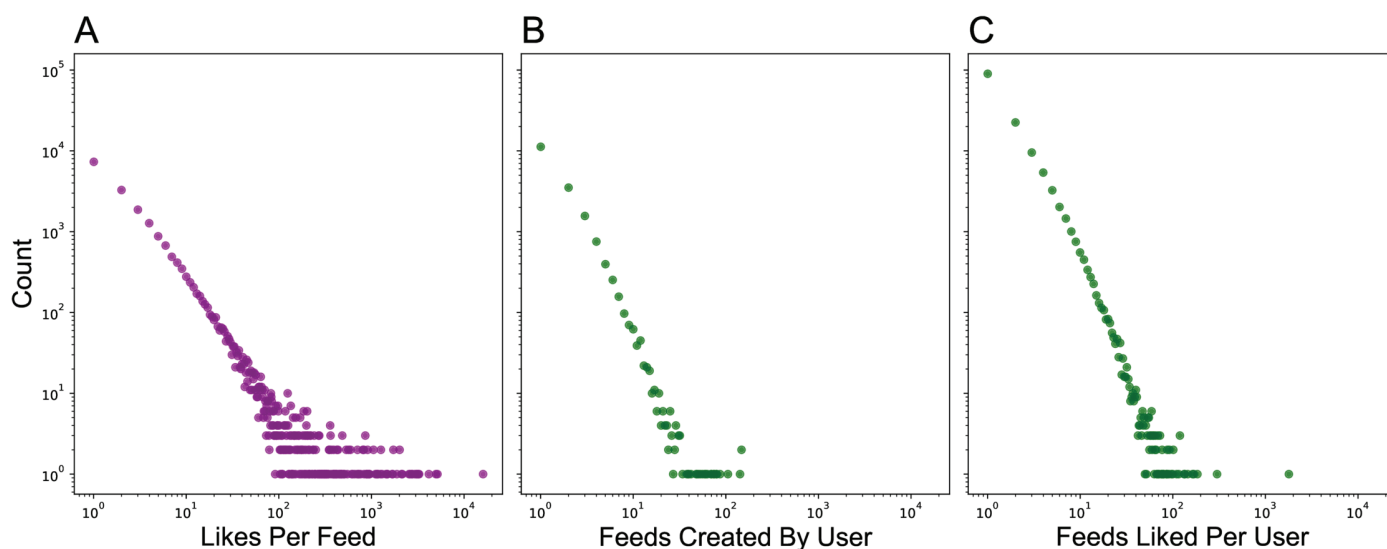


Fig 6. Log-log plots displaying distribution of feed related statistics. (A) Number of likes received per feed. (B) Number of feeds created per user. (C) Number of feeds liked per user.

<https://doi.org/10.1371/journal.pone.0318034.g006>

Table 4. Activity related to feeds on Bluesky. Included metrics are the mean μ , standard deviation σ , skewness γ , kurtosis β , m , maximum M , and the ratio of mean to maximum $\frac{\mu}{M}$ from the distributions depicted in Fig 2 and 3. α indicates the exponent of the power law distribution. The notably high values of γ and β indicate a pronounced right-skewed, heavy-tailed nature across all distributions. Furthermore, the exceptionally low $\frac{\mu}{M}$ values further confirm the extensive tail behaviour characteristic of these distributions.

Distribution	μ	σ	γ	β	m	M	$\frac{\mu}{M} \cdot 10^3$	α
# Likes per Feed	2.128	5.932	203.296	60,656.913	1	1,799	1.183	3.413
# Feeds Created per Person	2.161	13.931	122.996	16,082.634	1	1,828	1.182	3.350
# Number of Feeds Liked Per User	14.783	156.641	62.968	5,800.619	1	16,132	0.916	1.877

<https://doi.org/10.1371/journal.pone.0318034.t004>

Reddit banned thousands of subreddits, attracting banned extreme communities [32]. Similarly, Gab was founded to attract alt-right users [15] and Parler became the home of “disaffected right-wing social media users” [8]. While Mastodon has elements of techno-libertarian leaning [33] and took active measures to distance themselves from right-wing users [34], there exists to our knowledge no study examining the political leanings of users of the platform. While the literature contains ample examples of small social media platforms launched because of content moderation on Twitter, which were perceived to be disproportionately targeting conservative viewpoints [15], there is, to our knowledge, no empirical investigation into a predominantly left-leaning small platform.

Twitter, now X, is a platform that has a substantial user base with diverse political ideologies, ranging from far-right to far-left, although the majority of the user base has been characterized as left-leaning/center [35]. As shown in section *Activity on Bluesky*, sign-ups to Bluesky have been driven by activity on Twitter and its new leadership under Elon Musk. Since the purchase of Twitter by Elon Musk and its subsequent rebranding as X (the everything app), several newspapers and academics have reported that the user base, moderation philosophy, and goals of the platform have shifted towards a more right-leaning approach [36–38]. The perception of a shift towards the right on Twitter and the correlation of news about Twitter with sign-ups to Bluesky lead us to expect Bluesky to be predominantly left-wing, consisting of users who left the platform in search of a new social media site that is closer to their ideology. However, issue and platform polarization are hard to predict and strongly influenced by first movers and path dependencies [39]. We investigate the political leaning of Bluesky by extracting the domain of all links shared on Bluesky over the entire observation period. Table 5 lists the most shared Non-Political, Political, and “Questionable-Source” domains based on ratings by Media Bias Fact Check (MBFC) [40]. We classify a website as “Political” if its domain has an associated MBFC rating. We also report overall domain counts. Lastly, we show all “Questionable-Source” websites, filtered to include only those categorized by MBFC as either spreading fake news, conspiracies, or propaganda.

The most frequent non-political domains mostly relate to other social media websites. To ensure that we correctly mapped all links to domains, we expanded all links associated with a list of link-shorteners [41]. YouTube, the domain with the highest number of shares, was linked to a total of 1.66 million times. The second most shared domain is Spotify.com. Other

Table 5. Top ten domains by frequency in the dataset, comparing the overall occurrences to those classified under political and questionable sources (restricted to fake news, propaganda, or conspiracy) categories, respectively. To filter automated accounts, we exclude posts from accounts with more than 10,000 posted URLs.

Overall		Political		Questionable Sources	
Source Domain	Count	Source Domain	Count	Source Domain	Count
youtube.com	1,665,744	theguardian.com	124,123	dailymail.co.uk	2,501
spotify.com	237,786	nytimes.com	82,963	mondoweiss.net	2,249
tenor.com	185,865	washingtonpost.com	33,388	foxnews.com	870
twitch.tv	185,268	tagesschau.de	29,454	newsbreak.com	430
theguardian.com	124,123	bbc.com	27,973	indiatimes.com	401
substack.com	86,116	cnn.com	26,838	thenationalnews.com	318
instagram.com	83,077	spiegel.de	23,757	moveon.org	279
nytimes.com	82,963	apnews.com	23,305	almayadeen.net	270
twitter.com	77,440	reuters.com	23,230	theepochtimes.com	262
openstreetmap.org	68,503	theverge.com	21,578	presstv.ir	244
Total:	8,409,741	Total:	1,582,455	Total:	11,984
% of Total:	(100%)	% of Total:	(18.81%)	% of Total:	(0.14%)

<https://doi.org/10.1371/journal.pone.0318034.t005>

frequently shared social media domains include Twitch (185,268), Twitter (77,440), Instagram (83,077), and Substack (86,116) links. Tenor is a Gif sharing website. Interestingly, two political domains are among the most shared domains. *The Guardian* and the *New York Times* are classified by MBFC as “left-center”. In total, 408,133 unique domains were posted 8.409 million links to Bluesky.

Our analysis excludes posts from accounts that shared more than 10,000 URLs ($n=48$ accounts), which primarily represent automated news aggregators and content syndication services. The top domains shared by the 48 accounts are globo.com (a Brazilian news network, 176K shares), osintukraine.com (a war documentation and fact-checking site, 94K shares), yahoo.co.jp (a Japanese web portal, 82K shares), 9to5mac.com (an Apple news site, 60K shares), and lemonde.fr (a French newspaper, 48K shares). While these accounts contributed 15.30% of all URLs in the dataset, their removal had minimal impact on the overall distribution of content sources, with political sources shifting from 20.70% to 18.81% and questionable sources remaining nearly constant (0.13% to 0.14%). Though only 1.29% of unique domains in our dataset have an MBFC rating, this coverage increases substantially to 21.77% when considering domains shared by multiple users. In other words, the small subset of websites that were news and political sources (those with MBFC ratings) were shared dramatically more often than non-rated websites across the platform.

With the exception of *Tagesschau* and *Reuters*, all political outlets in the top ten political outlet columns are classified by MBFC as “left-center”. The two exceptions are both classified as “center” (or least biased). Prior to analyzing the overall distribution of all political domains in the dataset, this already indicates the bias of the platform. In total, we observed 1,582,455 occurrences of political domains being spread on Bluesky, making up 18.81% of all posts that include links in the dataset.

Compared to the spread of political domains, there is very little information stemming from websites classified as questionable sources, being spread. The top questionable source domain is “dailymail.co.uk” with only 2,501 occurrences in the entire dataset. Less than a percentage of posts including links on Bluesky contain links to domains classified as spreading fake news, conspiracies, or propaganda. The top 150 users spreading questionable news-sources make up 50% of the total spread on the platform. In total only 3,704 users on Bluesky have ever posted a domain associated with fake news, conspiracies, or propaganda.

Most of the domains spread via Bluesky are non-political. Within the domains with an associated political bias, left-leaning, specifically left-center, dominates. All but two of the top ten most spread political domains have an associated rating of left-center. Fig 7 shows the overall distribution of political domains spread via Bluesky. The bar chart on the left shows the distribution of all political domains on the website.

Over 63.4% of all domains are classified as left-leaning. Around 18.8% of the domains are classified as center, and 7.9% of the domains are right-leaning. This shows that Bluesky is mostly politically homogeneous, with a majority of the domains shared having an ideology left of center. No domains are classified by MBFC as extreme left and only 0.16% of domains are classified as extreme right. The bar charts on the right of the plot disaggregate the overall distribution into the distribution of right- and left-leaning outlets. For both left- and right-leaning outlets, the less extreme (i.e., more central) outlets dominate (left: 63.4%, extreme-left: 0%; right: 7.74%, extreme-right: 0.16%).

To examine the political stance of users of Bluesky, as opposed to the political stance of domains shared, we average the bias scores of all domains shared by each user. Each news outlet is assigned a political leaning score ranging from left to extreme right (assigned scores from -32 to +32). Looking at the average score of domains shared per user, 75.3% of users

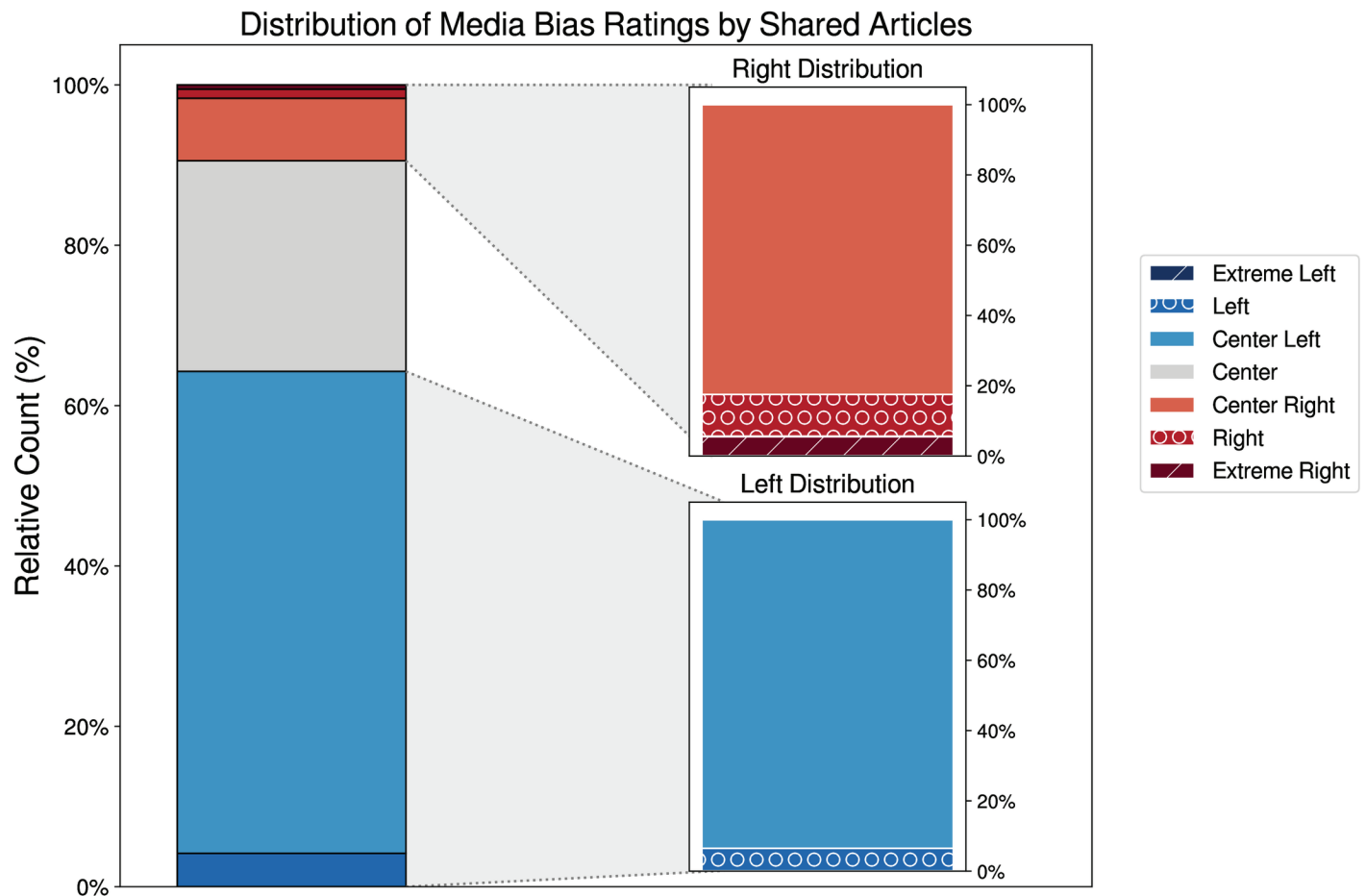


Fig 7. Distribution of political biases for domains posted on Bluesky, as categorized by Media Bias/Fact Check (MBFC). The main bar chart shows the overall distribution across six categories: left, center left, center, center right, right, and extreme right. Inset charts provide a detailed breakdown of left-leaning (left of center) and right-leaning (right of center) domains.

<https://doi.org/10.1371/journal.pone.0318034.g007>

are left of center (69.26% center-left, 6.05% left) and 4.81% of users are right of center (4.62% center right, 0.1% right, 0.08% extreme right). The remaining 19.79% are classified as center.

Cinelli et al. [42] study polarization and echo chambers by examining the distribution of user opinions and comparing them to the opinions of their neighborhood. The neighborhood is defined as the set of nodes directly connected to a given node in the network. The average opinion of the neighborhood is defined as $x_i^N \equiv \frac{1}{k_i^+} \sum_j A_{ij} x_j$ where k_i^+ is the out-degree of node i , A_{ij} is the adjacency matrix of the analysed network, and x_j is the opinion of neighbor j . The study uses the followership network as the basis of their analysis. We replicate their analysis of various social media sites, to investigate polarization on Bluesky, both with the interaction network of users (replying with a comment), and followership network.

Fig 8 displays the averaged political domain biases per user, as classified by MBFC, and juxtaposes them against the average biases of their neighborhood of users. Users are only included in the figure if they have shared at least five domains with a political bias and additionally have at least five neighbors who have done so too. This restricts our analysis to 43 thousand users. The distribution of user biases in the interaction and follower graph are generally very similar with a unimodal peak between center and left. This confirms our previous

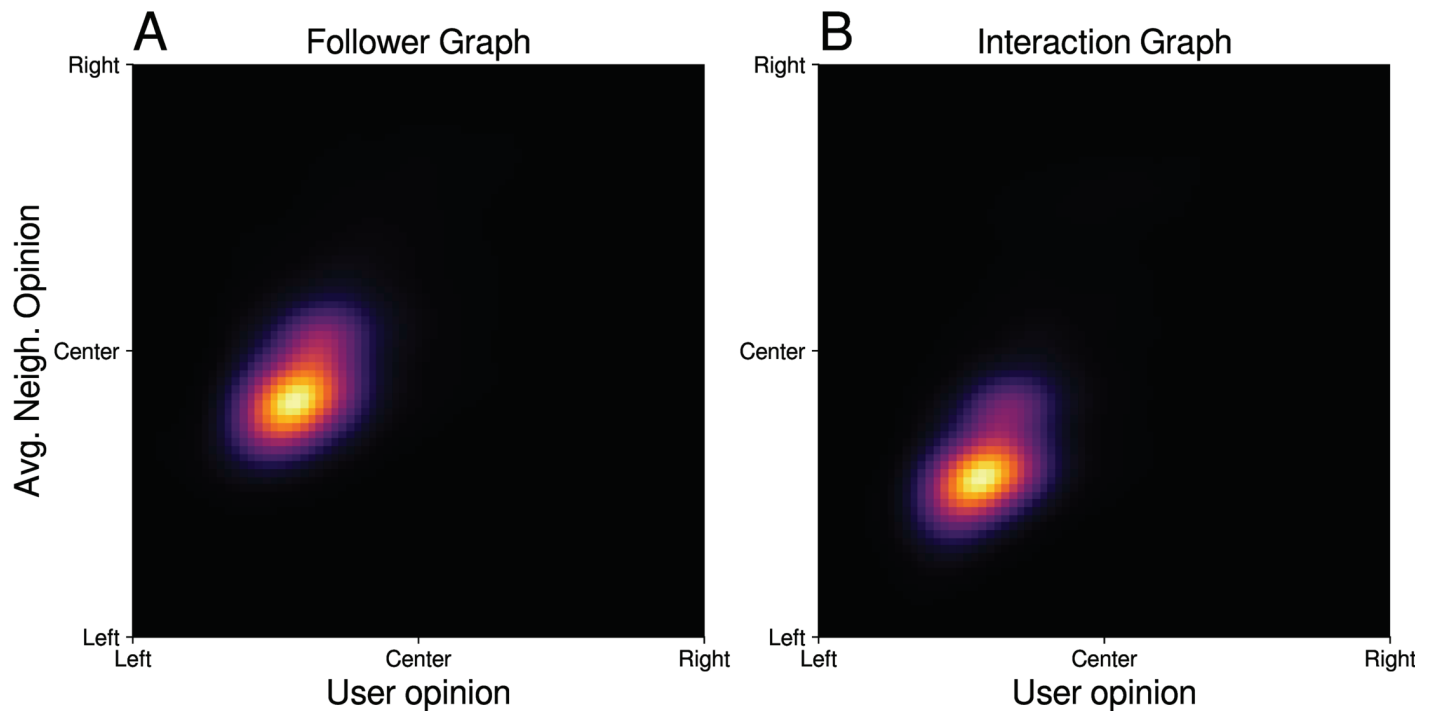


Fig 8. Heatmap of Political leaning of Users and the average of their neighborhood across the directed followership network (A) and the directed network of replies (B). Lighter areas indicate a higher density of users. Political leaning is calculated as the average political leaning of the URLs shared by a user. We exclude all users that have less than five neighbors or five posts. In total, 43,074 users have shared political domains at least five times and have at least five neighbors who have done so too and are thus included in the graphic

<https://doi.org/10.1371/journal.pone.0318034.g008>

findings that the vast majority of users on Bluesky have a center-left political leaning. Interestingly, the distribution of the average neighborhood bias of a users is more left leaning, less central, and more similar to the leaning of the seed user in the interaction graph than in the followership graph. This indicates that while users may follow a relatively more diverse range of political perspectives, their actual interactions are more politically homogeneous, highlighting a discrepancy between passive following behavior and active engagement patterns on the platform.

To investigate the polarization of opinions on a specific issue on Bluesky, we establish a corpus of posts related to the Israel-Palestine conflict and train a machine learning model to predict the stance of a post towards the conflict. Training details, test-set classification reports, and details on data labeling and querying are available in the Materials and Methods section of the manuscript.

Fig 9A shows the proportion of posts by stance per day. The y-axis represents the percentage of total posts for each stance, spanning from 0% to 100%. The x-axis covers the date range from July 2023 to early May 2024. The graph color-codes the posts: orange indicates neutral posts, green represents pro-Palestine posts, and blue signifies pro-Israel posts. The proportions of each stance change over time, with a notable dominance of neutral stances before October 7, 2023. On and after this date, there is a visible shift in the distribution of stances. Following the attacks on Israel, the percentage of neutral posts shrinks with an increase in both the number of Pro-Palestinian and Pro-Israel stances. Over the course of the following ten months, the percentage of Pro-Palestinian messages increases steadily, reaching the absolute majority of posts in January 2024.

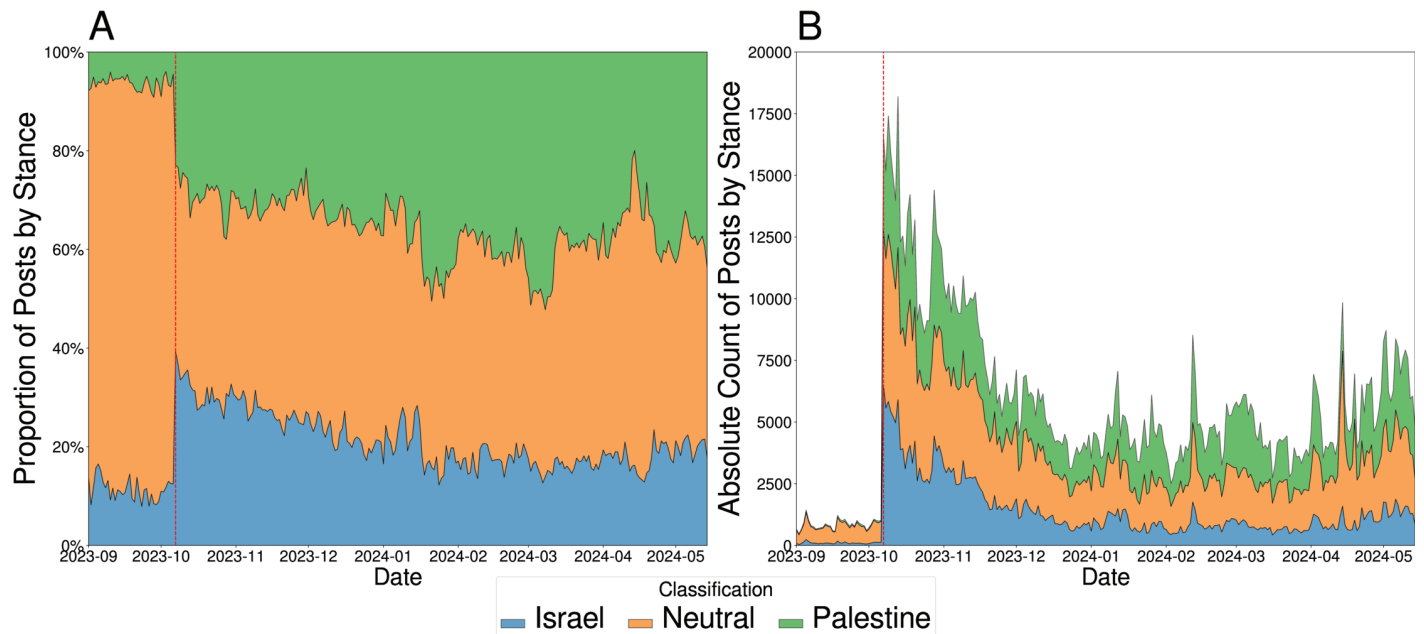


Fig 9. Distribution of posts by Stance on the Israel-Palestine conflict over time. (A) Daily proportions of posts, with the y-axis representing the percentage of total posts for each stance: neutral (orange), pro-Palestine (green), and pro-Israel (blue). Notably, neutral posts predominate until October 7, 2023, when a marked shift occurs towards more polarized views following the onset of the latest conflict. Over the subsequent months, pro-Palestine posts gradually outnumber pro-Israeli voices by January 2024. In the month preceding the attack on October 7th, the percentage of neutral posts dropped from 82.86% to 37.98%. While in October, pro-Israel voices outnumbered pro-Palestinian voices (33.01% vs 28.99%), in the final month of the observation period, only 20.74% of messages were pro-Israel, compared to 39.00% of messages containing a pro-Palestinian sentiment. (B) Absolute number of posts per day, with a significant spike in discussion beginning on October 7, 2023, followed by a stabilization in early 2024. This graph captures the fluctuations and trends in discourse surrounding the conflict from July 2023 to May 2024.

<https://doi.org/10.1371/journal.pone.0318034.g009>

Fig 9B displays the absolute count of posts by stance per day. Similar to Fig 9A, it uses the same color coding for each stance and spans the same time period on the x-axis. The y-axis, however, measures the count of posts, ranging from 0 to 18,000 posts per day. Prior to October 7, only a very small number of posts discussed Palestine & Gaza. On October 7, we see a spike with a gradual decay in posts until January 2024. Since then, the number of posts per day has remained relatively stable at around 4,000 messages.

To examine how users' stances on the Israel-Palestine conflict are distributed across the Bluesky network and how these stances relate to users' social connections, we conduct a network analysis similar to our earlier examination of general political ideology. This analysis allows us to visualize potential echo chambers or polarization specific to this issue, which may differ from the overall political leaning of the platform. We again extract all users with at least five posts indicating an opinion on the subject and average their political stances to map each user onto a one-dimensional stance. We then calculate the average stance of every user with at least five posts and five neighbors. The results are shown in Fig 10. Both networks showcase a similar neighborhood opinion graph with two distinct clusters. The majority of users are concentrated in two areas: a larger, more diffuse cluster spanning from Neutral to Pro-Palestine stances and a smaller, more compact cluster in the Pro-Israel region. A clear diagonal trend from bottom-left to top-right indicates users tend to connect with others holding similar views.

Although Bluesky predominantly displays a left-leaning political bias, this does not reflect uniformity of opinions on all subjects. The analysis of discussions on the Israel-Palestine

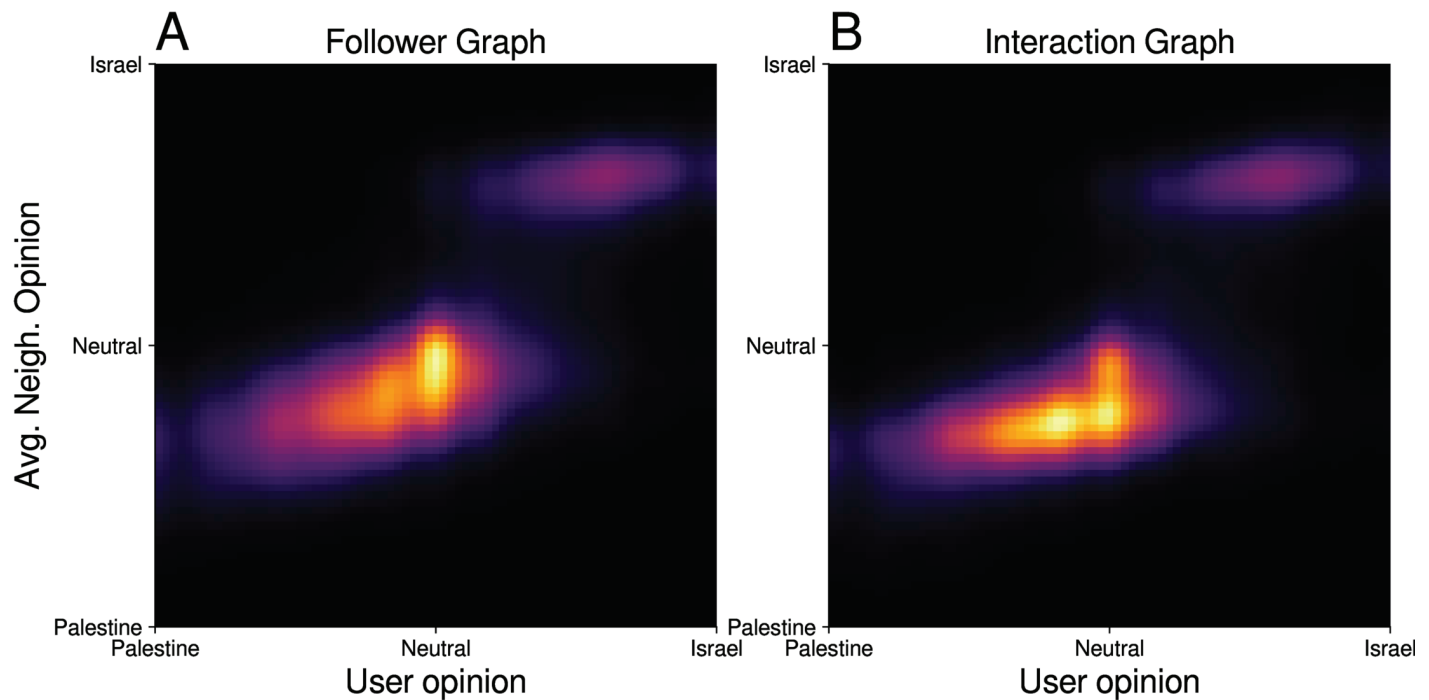


Fig 10. Heatmap of users' stance versus their neighborhood's average stance. The main plot shows the distribution across the directed followership network (A) and the directed network of replies (B). Lighter areas indicate a higher density of users. Political leaning is calculated as the average political leaning of URLs shared by a user. The figure includes 30,048 users who have shared at least five posts related to the conflict and have at least five neighbors who have done the same.

<https://doi.org/10.1371/journal.pone.0318034.g010>

conflict reveals a spectrum of stances within the platform. This range of perspectives highlights that political homogeneity does not necessarily dictate consensus on specific issues with polarized debates.

Discussion

Bluesky, for all its innovative features, is a social media site that resembles larger and older sites in almost all of its network features. Our analysis reveals patterns of clustering and small-world properties analogous to those observed in platforms like Twitter (now X). Our investigation into Bluesky's user composition reveals both homogeneity and diversity. While the platform exhibits a predominantly left-leaning user base in terms of broad political orientation, similar to some other small platforms, it demonstrates significant diversity in opinions on specific issues such as the Israel-Palestine conflict. Even within seemingly homogeneous platforms, there is potential for diverse discourse on particular topics. Future work could investigate whether such polarization of opinions on specific issues is indicative of a healthy dialogue in the marketplace of ideas or if it is driven by affective polarization and political sectarianism, which emerged in platforms such as Twitter and Facebook[43]. Our findings contribute to our understanding of how user bases form and evolve on emerging social media platforms and highlight the importance of investigating polarization by looking beyond a simple left-right spectrum. The creation of feeds has been taken up by users with enthusiasm with almost forty thousand choices present for users. However, only a small minority of users have liked a feed. Bluesky enables researchers to answer old questions with a novel treasure trove of data that could contribute to a range of scientific open questions.

Materials and methods

Method compliance statement

Data. The data for this study consists of the complete repositories of almost five million users on Bluesky, each containing all associated user profiles and actions. Due to the decentralized nature of the platform, Bluesky repository data is accessible to any person with the ID of a user. For each user, we first queried the centralized directory of user Decentralized IDs (DID PLC directory) [44] to get the personal data server address (PDS) of the user's repository. Given the Decentralized IDs (DID) of the user and the address of the PDS, the repository data can then be queried. Our dataset collection and analysis methods were designed to respect user privacy and platform terms. All data analyzed in this study consists only of public posts and user information accessed through standard API requests, in compliance with Bluesky's Terms of Service. To protect user privacy, we only report aggregated statistics and anonymized patterns. We did not collect any private messages, protected posts, or personally identifiable information beyond what users have made publicly available. To support reproducibility, we have made our analysis code, including repository download scripts, publicly available [45]. The repository also contains a text file of all 4,754,059 valid user DIDs present in the analysis. However, in accordance with privacy principles and to prevent potential re-identification, we cannot share the raw data collected for this study. This restriction has been imposed by PhF Ethics Committee at UZH, as the dataset contains potentially identifying information about users' social connections, posting patterns, and behavioral data that could be used to re-identify individuals even after de-identification. All data collection and analysis methods complied with Bluesky's Terms of Service and API usage guidelines. The data was collected using standard API endpoints available to all users, and no special access privileges were required or utilized. Our methods respected rate limits imposed by the platform.

An initial seed of 5.28 million IDs of Bluesky users was posted by a Bluesky contributor on the 26th of March 2024 [46]. Given this initial seed, we extracted all data for all users who remained active in the dataset and subsequently checked for any users referenced in the downloaded data. We repeated this procedure until no new users were found. The final dataset was cleaned and stored in a database containing individual tables for likes, follows, posts, reposts, blocks, and feed creations.

Table 6 contains a summary of the number of rows for each of the SQL tables. In addition, the column "unique users" contains the number of users who have authored an action (like, post, repost, follow, feed) in the table.

Interactions networks

Followership network. An edge connects a user to another if they follow that user. In contrast to Facebook, connections between users are not reciprocal, meaning that the network is directed. Followership relations are not transient but persistent. An edge between two users remains until it is removed. Following another user generally indicates an interest in the

Table 6. Number of rows and unique users per table in the dataset.

	Blocks	Follows	Posts	Likes	Reposts	Feeds
Total Rows	8,357,905	149,650,293	206,346,303	771,009,280	81,655,778	39,639
Unique Users	479,427	3,844,491	2,339,109	2,581,744	1,197,330	18,352

<https://doi.org/10.1371/journal.pone.0318034.t006>

content that they post, as their content will be shown to them in the user's main feed. Alternatively, followership could be an indicator for a social relationship outside of Bluesky, meaning that the two users are more likely to share socio-demographic features.

Replies network. An edge connects two users if one responds to another in the same thread with a comment. This network is also directed, but in contrast to the followership network, it is not persistent. Responding can indicate an overlap of the thematic interests of two users—but does not imply agreement.

Repost network. Two users are connected by an edge if a user reposts a post by another user. A repost on Bluesky is equivalent to a retweet on Twitter. The network is non-persistent and directed. A repost indicates an interest in the post of another user [47]. Additionally, the user is willing to share the content with their own followers [1].

Likes network. An edge in the Likes Network indicates whether a user liked a post from another user. The network is directed and non-persistent. Liking a post of another user indicates interest in the topics posted by the user [48]. In contrast to the Repost Network, the post will not be shown to the user's followers.

Stance detection for the Israel-Palestine conflict

Israel Palestine term extraction and data labelling. First, we extracted all posts containing the keywords “Israel”, “Palestine” and their translations into languages present in the dataset (Arabic, English, French, German, Greek, Hebrew, Italian, Japanese, Korean, Persian, Russian, Ukrainian, Azerbaijani, Danish, Dutch, Finnish, Hungarian, Indonesian, Kazakh, Chinese, Norwegian, Portuguese, Romanian, Slovene, Spanish, Swedish, Tajik, Turkish). Subsequently, we calculated the mutual information of all uni-, bi-, and tri-grams from the dataset with any of initial seed terms, filtering for a minimal number of occurrences of above 50 times. We then manually reviewed the top 100 n-grams, ranked by their mutual information, extracted for each of the initial seed terms and selected all n-grams directly related to the conflict. Lastly, to prevent any biases across languages, we translated all selected n-grams into all the initial languages to ensure that no variation in n-grams included bias the distribution of stance scores. A full list of all terms with at least 1,000 exclusive posts (posts not added by any other n-grams), can be found in [S1 Table](#) in the appendix. We then queried the database for any of the retrieved n-grams and created a dataset of 1.3 Million posts related to the conflict. From this subset, we manually annotated a random sample of 1,000 posts. Each post was labeled as Pro-Israel (1), Neutral (0), or Pro-Palestine (-1) based on the stance expressed in the content. An additional set of 1,000 posts was labeled by crowdworkers via [Appen.com](#). Crowdworkers were presented with posts from Bluesky and prompted with the following question. “Determine the stance (favor Israel, favor Palestine, or neither) of the author towards the Israel/Palestine conflict from each given social media post.” Each participant was given examples for each of the categories and additionally tested on ten quality assurance questions. The average agreement between annotators was 71.23%. We additionally removed any posts where no majority opinion among annotators emerged.

Stance prediction. Stance prediction involves automatically determining the position or attitude expressed in a piece of text toward a specific target or topic. We use the multilingual transformer-based language model XLM-RoBERTa large [49], which is well-suited for the stance prediction task due to its ability to capture cross-lingual semantic information. The data is preprocessed by tokenizing the text using the XLM-RoBERTa tokenizer, with a maximum sequence length of 128 tokens. The model is trained on an A100 GPU and evaluated on a hold-out test set. The final training configuration is reached via fine-tuning had a dropout-probability of 0.14, a learning rate of 2.2×10^{-05} and a weight decay of 0.001. In addition, we

freeze the first 21 (out of 24) layers of the model to reduce the chance of overfitting and accelerate training. The best model is finally determined by maximizing the Macro F1 score. Predictions are obtained by applying the trained model to the tokenized test data and selecting the class with the highest probability. Classification metrics, including precision, recall, and F1-score, are reported for each class in [table 7](#). The model significantly outperforms random guessing and a simple benchmarking naive bayes model which achieves a macro F1 score of 0.47 shown in [table 8](#).

Network metrics

Power law exponent. To estimate the exponent of the power-law degree distribution, we employed a maximum likelihood estimation method [50]. We considered only degrees $k \geq k_{min}$ where $k_{min} = 5$, and calculated the exponent α using the formula:

$$\alpha = 1 + n \left[\sum_{i=1}^n \ln \frac{k_i}{k_{min}} \right]^{-1}$$

where n is the number of nodes with degree $k \geq k_{min}$, and k_i are the observed degrees that meet this criterion.

Clustering coefficient. The clustering coefficient provides insights into the local structure of the network, indicating how likely it is for nodes to form tightly connected groups. A high clustering coefficient suggests a network with many triangles. In the context of social networks, the clustering coefficient can be intuitively understood as the probability that two of your friends are also friends with each other. More formally, it measures the likelihood of triadic closure in the network. The clustering coefficient for a directed graph is defined by considering the possible directed triangles[51].

For each node i , the local clustering coefficient C_i is computed as:

$$C_i = \frac{\text{number of directed triangles passing through } i}{\text{number of possible directed triangles passing through } i}$$

Table 7. Test set classification performance.

Class	Precision	Recall	F1-Score	Support
Palestine	0.623	0.716	0.667	60
Neutral	0.712	0.602	0.653	78
Israel	0.595	0.647	0.620	34
Accuracy				0.653
F1 Score				0.646

<https://doi.org/10.1371/journal.pone.0318034.t007>

Table 8. Test set classification report for naive Bayes.

Class	Precision	Recall	F1-score	Support
Palestine	0.471	0.817	0.598	60
Neutral	0.565	0.333	0.419	78
Israel	0.500	0.324	0.393	34
Accuracy				0.500
F1 Score				0.470

<https://doi.org/10.1371/journal.pone.0318034.t008>

Which is calculated as:

$$C_i = \frac{(A + A^T)_{ii}^3}{2(d_i^{\text{tot}}(d_i^{\text{tot}} - 1) - 2d_i^{\leftrightarrow})}$$

where $(A + A^T)_{ii}^3$ is the number of directed triangles including node i , d_i^{tot} is the sum of in and out degrees of node i . Lastly, d_i^{\leftrightarrow} is the reciprocal degree of node i , i.e. the number of nodes j for which both an edge $i \rightarrow j$ and an edge $j \rightarrow i$ exist. For the entire network, the clustering coefficient C is computed as the average of all individual local clustering coefficients. We normalize the clustering coefficient by creating a randomized configuration-model graph with the same degree in- and out-degree sequences than the empirical graphs. In the random graph, the in- and out-degree of each node is pre-defined, but nodes are randomly connected. The normalized clustering coefficient is defined as $CC_{\text{norm}} = \frac{CC}{CC_{\text{rand}}}$ [52].

Network density. Network density measures the proportion of potential connections in a network that are actual connections. It provides insight into the overall connectivity and compactness of the graph. For a directed graph, the network density D is defined as:

$$D = \frac{m}{n(n-1)}$$

where m is the total number of directed edges in the graph, and n is the total number of nodes.

Average shortest path length. The average shortest path length is a measure of the efficiency of information or traffic flow within a network. It quantifies the average number of steps along the shortest paths for all possible pairs of network nodes. It is a significant indicator of the 'small-world' characteristic of a network. A network exhibits the 'small-world' characteristic if "any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances" [53] [52]. This metric indicates the ease with which information spreads across the network and is a key factor in the analysis of network efficiency and connectivity. We provide an approximate value for the average shortest path in each network by randomly sampling 50 thousand pairs of nodes in the graph and calculating the mean distance between the nodes.

Degree distribution. The degree distribution of a network describes the relative frequency of nodes with different degrees within the graph. In social media networks, this distribution is often heavy-tailed, with a power-law-like shape, indicating that while most users have few connections, a small number of users (hubs) have a disproportionately large number of connections, but also the absence of clear scale separations between users based on connections [54] [55].

Supporting information

S1 Text: Appendix.
(PDF)

Acknowledgements

The authors would like to thank Ilya (Marshal) Siamionau for his technical support with the Bluesky API. We also thank Frederic Denker and Matteo Cinelli for their valuable inputs, as well as our colleagues Yasaman Asgari, Samuel Koovely, and Yuan Zhang.

Author contributions

Conceptualization: Dorian Quelle, Alexandre Bovet.

Data curation: Dorian Quelle.

Formal analysis: Dorian Quelle.

Investigation: Dorian Quelle.

Methodology: Dorian Quelle, Alexandre Bovet.

Project administration: Alexandre Bovet.

Resources: Alexandre Bovet.

Supervision: Alexandre Bovet.

Validation: Dorian Quelle, Alexandre Bovet.

Visualization: Dorian Quelle.

Writing – original draft: Dorian Quelle.

Writing – review & editing: Dorian Quelle, Alexandre Bovet.

References

1. Kleppmann M, Frazee P, Gold J, Graber J, Holmgren D, Ivy D, et al. Bluesky and the AT Protocol: Usable Decentralized Social Media. arXiv preprint arXiv:240203239. 2024;.
2. Dorsey J. Twitter is funding a small independent team of up to five open source architects...; 2019. <https://x.com/jack/status/1204766078468911106>.
3. Jaz. Statistics Overview on Bsky; 2023. <https://bsky.jazco.dev/stats>.
4. Graysky. Graysky: Now Available!; 2023. <https://graysky.app>.
5. Deck Blue. Title of the Specific Content; 2023. <https://deck.blue>.
6. Failla A, Rossetti G. "I'm in the Bluesky Tonight": Insights from a Year Worth of Social Data. arXiv preprint arXiv:240418984. 2024.
7. Zignani M, Gaito S, Rossi GP. Follow the "Mastodon": Structure and Evolution of a Decentralized Online Social Network. ICWSM. 2018;12(1):541–50. <https://doi.org/10.1609/icwsm.v12i1.14988>
8. Aliapoulos M, Bevensee E, Blackburn J, Bradlyn B, De Cristofaro E, Stringhini G, et al. An early look at the Parler online social network. arXiv. Preprint posted online on January. 2021;11.
9. Bluesky. Bluesky - What's Hot; 2024. Available from: <https://bsky.app/profile/bsky.app/feed/whats-hot>.
10. Kubin E, von Sikorski C. The role of (social) media in political polarization: a systematic review. Ann Int Commun Assoc. 2021;45(3):188–206. <https://doi.org/10.1080/23808985.2021.1976070>
11. Stocking G, Mitchell A, Matsa KE, Widjaya R, Jurkowitz M, Ghosh S, et al. The role of alternative social media in the news and information environment. Pew Research Center. 2022.
12. Flamino J, Galeazzi A, Feldman S, Macy MW, Cross B, Zhou Z, et al. Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. Nat Hum Behav. 2023;7(6):904–16. <https://doi.org/10.1038/s41562-023-01550-8> PMID: 36914806
13. González-Bailón S, Lazer D, Barberá P, Zhang M, Allcott H, Brown T, et al. Asymmetric ideological segregation in exposure to political news on Facebook. Science. 2023;381(6656):392–8. <https://doi.org/10.1126/science.ade7138> PMID: 37499003
14. Gerard P, Botzer N, Weninger T. Truth social dataset. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 17; 2023. p. 1034–40.
15. Sharevski F, Jachim P, Pieroni E, Devine A. "Gettr-ing" Deep Insights from the Social Network Gettr. arXiv preprint arXiv:220404066. 2022.
16. Than N, Yoong D, Rodriguez MY, Windel FM. "Welcome to Gab": Exploring Political Discourses in a Non-Moderated Social Media Platform. IDEAH. 2021;2(1).
17. Jeong U, Nirmal A, Jha K, Tang SX, Bernard HR, Liu H. User migration across multiple social media platforms. In: Proceedings of the 2024 SIAM International Conference on Data Mining (SDM). SIAM; 2024. p. 436–444.

18. Guardian T. Elon Musk says Twitter, now X, could charge all users subscription fees. *The Guardian*. 2023.
19. BBC. Elon Musk: Social media platform X, formerly Twitter, could go behind paywall. *BBC News*. 2023.
20. Vigliarolo B. Twitter rate-limits itself into a weekend of chaos. *The Register*. 2023.
21. Bloomberg. Twitter's Troubles Are Perfectly Timed for Meta. *Bloomberg*. 2024.
22. Guardian TX, formerly Twitter, rolls out US\$1 annual fee for new users in New Zealand and the Philippines. *The Guardian*. 2023.
23. Support T. Starting today, we're testing a new program (Not A Bot) in New Zealand and the Philippines. New, unverified accounts will be required to sign up for a \$1 annual subscription to be able to post & interact with other posts. 2023. Available from: <https://twitter.com/Support/status/1714429406192582896>.
24. Euronews. X, formerly Twitter, sees massive outage as tens of thousands report issues. *Euronews*. 2023.
25. Independent. Is X/Twitter down? Users report problems accessing feeds in multiple countries. *The Independent*. 2023.
26. Dickison ME, Magnani M, Rossi L. Multilayer social networks. Cambridge University Press; 2016.
27. Magnani M, Rossi L. The ml-model for multi-layer social networks. In: 2011 International conference on advances in social networks analysis and mining. IEEE; 2011. p. 5–12.
28. Jeong U, Jiang B, Tan Z, Bernard HR, Liu H. BlueTempNet: A Temporal Multi-network Dataset of Social Interactions in Bluesky Social. *arXiv preprint arXiv:240717451*. 2024.
29. Bluesky. Bluesky - For Science; 2024. Available from: <https://bsky.app/profile/did:plc:jfhpnst6flqway4eaeqzj2a/feed/for-science>.
30. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:220305794*. 2022.
31. Reimers N, Gurevych I. Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:200409813*. 2020.
32. Mekacher A, Papasavva A. "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. vol. 16; 2022. p. 1302–11.
33. Nicholson MN, Keegan BC, Fiesler C. Mastodon rules: characterizing formal rules on popular Mastodon instances. In: Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing; 2023. p. 86–90.
34. Gehl RW, Zulli D. The digital covenant: non-centralized platform governance on the mastodon social network. *Inform Commun Soc*. 2022;26(16):3275–91. <https://doi.org/10.1080/1369118x.2022.2147400>
35. Wojcik S, Hughes A. Sizing up Twitter users. *PEW research center*. 2019;24:1–23.
36. Pew Research Center. After Musk's takeover, big shifts in how Republican and Democratic Twitter users view the platform; 2023. Available from: <https://www.pewresearch.org/short-reads/2023/05/01/after-musks-takeover-big-shifts-in-how-republican-and-democratic-twitter-users-view-the-platform/#:~:text=Musk%20has%20been%20a%20vocal,going%20unchecked%20on%20the%20site>.
37. Economist T. Has Twitter, now X, become more right-wing?; 2023. Available from: <https://www.economist.com/graphic-detail/2023/12/20/has-twitter-now-x-become-more-right-wing>.
38. Guardian T. Twitter amplifies conservative media under Elon Musk, data shows; 2023. Available from: <https://www.theguardian.com/world/2023/jun/03/twitter-conservative-media-elon-musk-ron-desantis>.
39. Macy M, Deri S, Ruch A, Tong N. Opinion cascades and the unpredictability of partisan polarization. *Sci Adv*. 2019;5(8):eaax0754. <https://doi.org/10.1126/sciadv.aax0754> PMID: 31489373
40. Zandt DMV. Media Bias/Fact Check; 2024. <https://drive.google.com/file/d/0B0T02oNLikiuU0s2NIdaWmNZbUk/view?pli=1>
41. Kai S. ShortURL Services List; 2023. <https://github.com/sambokai/ShortURL-Services-List>
42. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. *Proc Natl Acad Sci U S A*. 2021;118(9):e2023301118. <https://doi.org/10.1073/pnas.2023301118> PMID: 33622786
43. Finkel EJ, Bail CA, Cikara M, Ditto PH, Iyengar S, Klar S, et al. Political sectarianism in America. *Science*. 2020;370(6516):533–6. <https://doi.org/10.1126/science.abe1715>.
44. Bluesky PBC. DID PLC Directory; 2023. <https://web.plc.directory>.

45. Quelle D, Bovet A. Code To Reproduce the Analysis of “Bluesky Network Topology, Polarization, and Algorithmic Curation”; 2024. Available from: <https://doi.org/10.7910/DVN/NGQKDS>
46. Jaz. Bluesky Post; 2024. <https://bsky.app/profile/jaz.bsky.social/post/3konq6ph3nn23>.
47. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: 2010 43rd Hawaii international conference on system sciences. IEEE; 2010. p. 1–10.
48. Levordashka A, Utz S, Ambros R. What’s in a Like? Motivations for Pressing the Like Button. ICWSM. 2021;10(1):623–6. <https://doi.org/10.1609/icwsm.v10i1.14768>
49. Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, et al. Unsupervised cross-lingual representation learning at scale. CoRR. 2019;abs/1911.02116.
50. Newman M. Power laws, Pareto distributions and Zipf’s law. Contemp Phys. 2005;46(5):323–51. <https://doi.org/10.1080/00107510500052444>
51. Fagiolo G. Clustering in complex directed networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007;76(2 Pt 2):026107. <https://doi.org/10.1103/PhysRevE.76.026107> PMID: 17930104
52. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. Nature. 1998;393(6684):440–2. <https://doi.org/10.1038/30918> PMID: 9623998
53. Kleinberg J. The small-world phenomenon: An algorithmic perspective. In: Proceedings of the thirty-second annual ACM symposium on Theory of computing; 2000. p. 163–70.
54. Raban DR, Rabin E. Statistical inference from power law distributed web-based social interactions. Internet Res. 2009;19(3):266–78. <https://doi.org/10.1108/10662240910965342>
55. Centola D. The Social Origins of Networks and Diffusion. Am J Sociol. 2015;120(5):1295–338. <https://doi.org/10.1086/681275> PMID: 26421341