Do Servers Matter on Mastodon? Data-driven Design for Decentralized Social Media

Carl Colglazier[1]

[1]Northwestern University
carlcolglazier@u.northwestern.edu

Abstract

When trying to join Mastodon, a decentralized collection of interoperable social networking servers, new users face the dilemma of choosing a home server. Using trace data from millions of new Mastodon accounts, we show that new accounts are less likely to remain active on the network's largest general instances compared to others. Additionally, we observe a trend of users migrating from larger to smaller servers. Addressing the challenge of onboarding and server selection, the paper proposes a decentralized recommendation system for server using hashtags and the Okapi BM25 algorithm. This system leverages servers' top hashtags and their frequency to create a recommendation mechanism that respects Mastodon's decentralized ethos.

Introduction

Following Twitter's 2022 acquisition, Mastodon—an open-source, decentralized social network and microblogging community—saw an increase in activity and attention as a potential Twitter alternative (He et al. 2023; La Cava, Aiello, and Tagarelli 2023). While millions of new accounts significantly increased the size of the network, many newcomers found the process confusing and did not remain active. Unlike centralized social media platforms, Mastodon is a network of independent servers, each with their own rules and norms (Nicholson, Keegan, and Fiesler 2023), which can communicate with each other using the shared ActivityPub protocols. Athough accounts can move between Mastodon servers, the local experience can vary widely from server to server.

Attracting and retaining newcomers is a key challenge for online communities (Kraut, Resnick, and Kiesler 2011, 182). On Mastodon, the onboarding process has not always been straightforward: variation among servers mean newcomers who may not even be aware of the specific rules, norms, or general topics of interest on the server they are joining (Diaz 2022). Various guides and resources for people trying to join Mastodon offered mixed advice on choosing a server. Some suggest that the most important thing is to simply join any server and work from there (Krasnoff 2022; Silberling 2023); others have created tools and guides to help people find potential servers of interest by size and location(TheKinrar 2017; King 2024).

Mastodon's decentralized design has long been in tension with the disproportionate popularity of a small set of large, general-topic servers within the system (Raman et al. 2019). Analysing the activity of new accounts that join the network, we find that users who sign up on such servers are less likely to remain active after 91 days. We also find that users who move accounts tend to gravitate toward smaller, more niche servers over time, suggesting that established users may also find additional utility from such servers.

In response to these findings, we propose a potential extension to Mastodon to facilitate server and tag recommendations by having each server report their most popular local hashtags. This recommendation system could both help newcomers find servers that match their interests and help established accounts discover "neighborhoods" of related servers.

Background

Empirical Setting

The Fediverse is a set of decentralized online social networks which interoperate using shared protocols like ActivityPub. Mastodon is a software program used by many Fediverse servers and offers a user experience similar to the Tweetdeck client for Twitter. It was first created in late 2016 and saw a surge in interest in 2022 during and after Elon Musk's Twitter acquisition.

Mastodon features three kinds of timelines: a "home" timeline which shows all posts from accounts followed by the user; a "local" timeline which shows all public posts from the local server; and a "federated" timeline which includes all posts from users followed by other users on their server. The local timeline is unique to each server. On larger servers, this timeline can be unwieldy; however, on smaller servers, it presents the opportunity to discover new posts and users of potential interest.

Discovery has been challenging on Mastodon. Text search, for instance, was impossible on most servers until support for this feature was added on an optional, opt-in basis using Elasticsearch in late 2023 (Rochko 2023a). Recommendation systems are currently a somewhat novel problem in the context of decentralized online social net-

works. Trienes, Cano, and Hiemstra (2018) developed a recommendation system for finding new accounts to follow on the Fediverse which used collaborative filtering based on BM25 in an early example of a content discovery system on Mastodon.

Individual Mastodon servers can have an effect on the end experience of users. For example, some servers may choose to federate with some servers but not others, altering the topology of the Fediverse network for their users. At the same time, accounts need to be locked into one specific server. Because of Mastodon's data portability, users can move their accounts freely between servers while retaining their followers, though their post history remains with their original account.

The Mastodon Migrations

Mastodon saw a surge in interest in 2022 and 2023, particularly after Elon Musk's Twitter acquisition. In particular, four events of interests drove measurable increases in new users to the network: the announcement of the acquisition (April 14, 2022), the closing of the acquisition (October 27, 2022), a day when Twitter suspended a number of prominent journalists (December 15, 2022), and a day when Twitter experienced an outage and started rate limiting accounts (July 1, 2023). Many Twitter accounts announced they were setting up Mastodon accounts and linked their new accounts to their followers, often using tags like #TwitterMigration (He et al. 2023) and driving interest in Mastodon in a process La Cava, Aiello, and Tagarelli (2023) found consistent with social influence theory.

Despite the influx of users, not all of these new accounts remained active. As such, some media outlets have framed reports on Mastodon (Hoover 2023) through what Zulli, Liu, and Gehl (2020) calls the "Killer Hype Cycle", whereby the media finds a new alternative social media platform, declares it a potential killer of some established platform, and later calls it a failure if it does not displace the existing platform. Such framing fails to take systems like the Fediverse seriously for their own merits: completely replacing existing commercial systems is not the only way to measure success, nor does it account for the real value the Fediverse provides for its millions of active users.

Mastodon's approach to onboarding has changed over time. In much of 2020 and early 2021, the Mastodon developers closed sign-ups to their flagship server and linked to an alternative server, which saw increased sign-ups during this period. They also linked to a list of servers on the "Join Mastodon" webpage[1], where all servers are pre-approved and follow the Mastodon Server Covenant which guarantees certain content moderation standards and data protections. Starting in 2023, the Mastodon developers shifted toward making the flagship server the default when people sign up on the official Mastodon Android and iOS apps (Rochko 2023b; Roth 2023). These changes suggest that removing friction to onboarding is an increasing priority for the Mastodon developers.

[1]https://joinmastodon.org/servers

Newcomers in Online Communities

Onboarding newcomers is an important part of the life cycle of online communities. Any community can expect a certain amount of turnover, and so it is important for the long-term health and longevity of the community to be able to bring in new members (Kraut, Resnick, and Kiesler 2011, 182). However, the process of onboarding newcomers is not always straightforward.

The series of migrations of new users into Mastodon in many ways reflect folk stories of "Eternal Septembers" on previous communication networks, where a large influx of newcomers challenged the existing norms (Driscoll 2023; Kiene, Monroy-Hernández, and Hill 2016). Many Mastodon servers do have specific norms which people coming from Twitter may find confusing, such as local norms around content warnings (Nicholson, Keegan, and Fiesler 2023). Variation among servers can also present a challenge for newcomers who may not even be aware of the specific rules, norms, or general topics of interest on the server they are joining (Diaz 2022). Mastodon servers open to new accounts must thus be both accommodating to newcomers while at the same ensuring the propagation of their norms and culture, either through social norms or through technical means.

Data

Mastodon has an extensive API which allows for the collection of public posts and account information. We collected data from the public timelines of Mastodon servers using the Mastodon API with a crawler which runs once per day. We also collected account information from the opt-in public profile directories on these servers.

Mastodon Profiles: We collected accounts using data previously collected from posts on public Mastodon timelines from October 2020 to August 2023. We then queried for up-to-date information on those accounts including their most recent status and if the account had moved as of February 2024. Through this process, we discovered a total of 1,784,438 account created between August 14, 2020 and January 1, 2024. We then filtered out accounts which were bots (19,014 accounts), had been suspended (163,721 accounts), had been marked as moved to another account (41,111 accounts), had been limited by their local server (6,062 accounts), had no statuses (140,870 accounts), or had posted their last status on the same day as their account creation (391,173 accounts). This gave us a total of 1,067,546 accounts which met all the filtering criteria. Note that because we got updated information on each account, we include only accounts on servers which still existed at the time of our profile queries and which returned records for the account.

Tags: Mastodon supports hashtags, which are user-generated metadata tags that can be added to posts. Clicking the link for a tag shows a stream of posts which also have that tag from the federated timeline, which includes accounts on the same server and posts from accounts followed by the accounts on the local server. We collected 1,206,779 statuses posted by 67,604 accounts on 723 unique servers from between May to July 2023 which contained at least

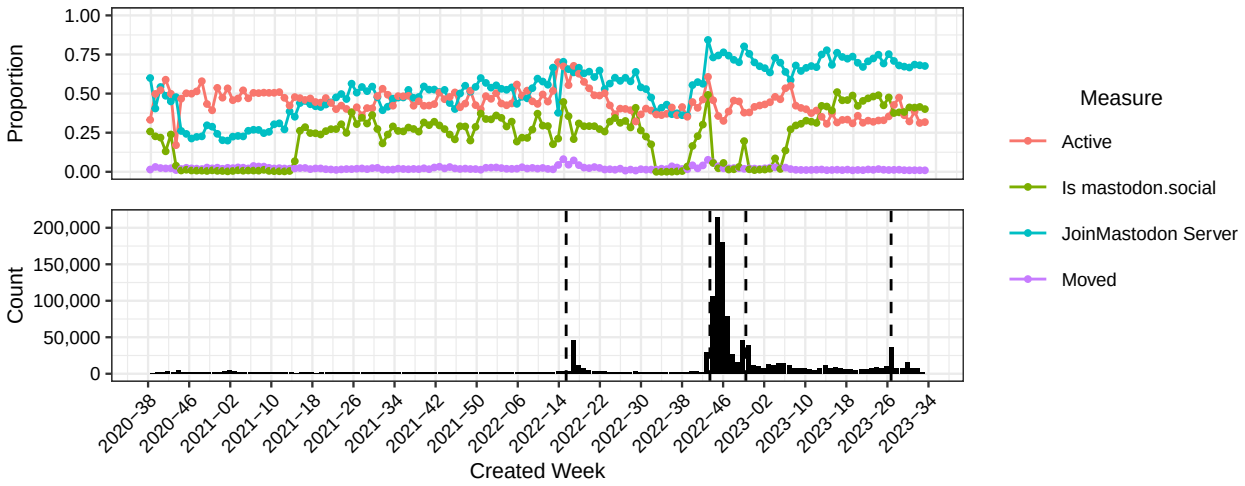

Figure 1: Accounts in the dataset created between January 2022 and March 2023. The top panels shows the proportion of accounts still active 45 days after creation, the proportion of accounts that have moved, and the proportion of accounts that have been suspended. The bottom panel shows the count of accounts created each week. The dashed vertical lines in the bottom panel represent the annoucement day of the Elon Musk Twitter acquisition, the acquisition closing day, a day where Twitter suspended a number of prominent journalist, and a day when Twitter experienced an outage and started rate limiting accounts.

one hashtag.

Analysis and Results

Survival Model

Are accounts on suggested general servers less likely to remain active than accounts on other servers?

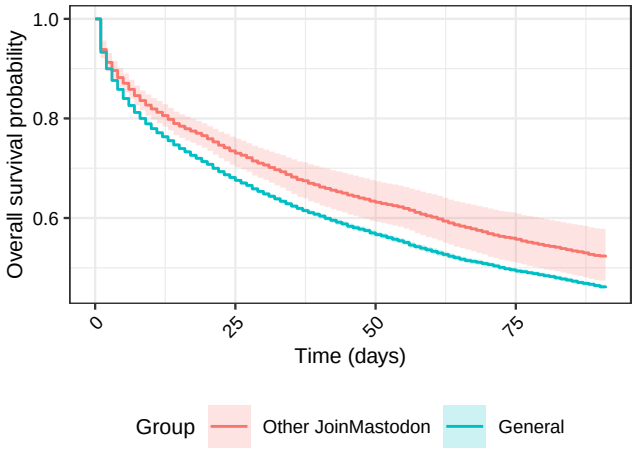

Figure 2: Survival probabilities for accounts created during May 2023.

Using 28,258 accounts created from May 1 to June 30, 2023, we create a Kaplan–Meier estimator for the probability that an account will remain active based on whether the account is on one of the largest general instances[2] featured at the top of the Join Mastodon webpage or otherwise if it is on a server in the Join Mastodon list. Accounts are considered active if they have made at least one post after the censorship period 91 days after account creation.

We find that accounts on the largest general instances are less likely to remain active than accounts on other servers, while accounts created on smaller servers are more likely to remain active.

[2]mastodon.social, mstdn.social, mastodon.world, mas.to, mastodon.online, mastodonapp.uk, universeodon.com, masto.ai, c.im, social.vivaldi.net, mstdn.party, ohai.social

Table 1: Exponential family random graph models for account movement between Mastodon servers. Accounts were created in May 2022 and moved to another account at some later point.

	Coef.	Std.Error
Sum	−9.529	***0.188
Nonzero	−3.577	***0.083
Smaller server	0.709	***0.032
Server size (outgoing)	0.686	***0.013
Open registrations (incoming)	0.168	***0.046
Languages match	0.044	0.065

Moved Accounts

Do accounts tend to move to larger or smaller servers?

Mastodon users can move their accounts to another server while retaining their connections (but not their posts) to other Mastodon accounts. This feature, built into the Mastodon software, offers data portability and helps avoid lock-in.

To corroborate our findings, we also use data from thousands of accounts which moved between Mastodon servers, taking advantage of the data portability of the platform. Conceiving of these moved accounts as edges within a weighted

directional network where nodes represent servers, edges represent accounts, and weights represent the number of accounts that moved between servers, we construct an exponential family random graph model (ERGM) with terms for server size, open registrations, and language match between servers. We find that accounts are more likely to move from larger servers to smaller servers.

Proposed Recommendation System

How can we build an opt-in, low-resource recommendation system for finding Fediverse servers?

Based on these findings, we suggest a need for better ways for newcomers to find servers and propose a viable way to create server and tag recommendations on Mastodon. This system could both help newcomers find servers that match their interests and help established accounts discover "neighborhoods" of related servers.

One challenge in building such a system is the decentralized nature of the system. A single, central actor which collects data from servers and then distributes recommendations would be antithetical to the decentralized nature of Mastodon. Instead, we propose a system where servers can report the top hashtags by the number of unique accounts on the server using them during the last three months. Such a system would be opt-in and require few additional server resources since tags already have their own database table.

Recommendation System Design

We use Okapi BM25 to construct a term frequency-inverse document frequency (TF-IDF) model to associate the top tags with each server using counts of tag-account pairs from each server for the term frequency and the number of servers that use each tag for the inverse document frequency. We then L2 normalize the vectors for each tag and calculate the cosine similarity between the tag vectors for each server.

$$tf = \frac{f_{t,s} \cdot (k_1 + 1)}{f_{t,s} + k_1 \cdot (1 - b + b \cdot \frac{|s|}{avgstl})}$$

where $f_{t,s}$ is the number of accounts using the tag t on server d, k_1 and b are tuning parameters, and $avgstl$ is the average sum of account-tag pairs. For the inverse document frequency, we use the following formula:

$$idf = \log \frac{N - n + 0.5}{n + 0.5}$$

where N is the total number of servers and n is the number of servers where the tag appears as one of the top tags. We then apply L2 normalization:

$$tfidf = \frac{tf \cdot idf}{\|tf \cdot idf\|_2}$$

Applications

Server Similarity Neighborhoods Mastodon provides two feeds in addition to a user's home timeline populated by accounts they follow: a local timeline with all public posts from their local server and a federated timeline which includes all posts from users followed by other users on their server. We suggest a third kind of timeline, a *neighborhood timeline*, which filters the federated timeline by topic. We calculate the pairwise similarity between two servers using cosine similarity.

Tag Similarity We also calculate the similarity between tags using the same method. This can be used to suggest related tags to users based on their interests or tags related to already selected tags in the recommendation system.

Server Discovery Given a set of popular tags and a list of servers, we build a recommendation system[3] where users select tags from a list of popular tags and receive server suggestions. The system first creates a subset of vectors based on the TF-IDF matrix which represents the top clusters of topics. After a user selects the top tags of interest to them, it suggests servers which match their preferences using the singular value decomposition (SVD) of the TF-IDF matrix.

Discussion

The analysis can also be improved by additionally focusing on factors lead to accounts remaining active or dropping out, which a particular focus on the actual activity of accounts over time. For instance, do accounts that interact with other users more remain active longer? Are there particular markers of activity that are more predictive of account retention? Future work could use these to provide suggests for ways to helps newcomers during the onboarding process.

The observational nature of the data limit some of the causal claims we can make. It is unclear, for instance, if accounts on general servers are less likely to remain active because of the server itself or because of the type of users who join such servers. For example, it is conceivable that the kind of person who spends more time researching which server to join is more invested in their Mastodon experience than one who simply joins the first server they find.

Future work is necessary to determine the how well the recommendation system is at helping users find servers that match their interests. This may involve user studies and interviews to determine how well the system works in practice.

While the work presented here is based on observed posts on the public timelines, simulations may be helpful in determining the robustness of the system to targeted attacks. Due to the decentralized nature of the system, it is feasible that a bad actor could set up zombie accounts on servers to manipulate the recommendation system. Simulations could help determine how well the system can resist such attacks and ways to mitigate this risk.

Conclusion

Based on analysis of trace data from millions of new Fediverse accounts, we find evidence that suggests that servers matter and that users tend to move from larger servers to smaller servers. We then propose a recommendation system that can help new Fediverse users find servers with a high probability of being a good match based on their interests.

[3] A live demo for the system is availible at https://carlcolglazier.com/demos/deweb2024/

Based on simulations, we demonstrate that such a tool can be effectively deployed in a federated manner, even with limited data on each local server.

References

Diaz, J. 2022. Using Mastodon Is Way Too Complicated to Ever Topple Twitter. https://www.fastcompany.com/90808984/using-mastodon-is-way-too-complicated-to-ever-topple-twitter.

Driscoll, K. 2023. Do We Misremember Eternal September?

He, J.; Zia, H. B.; Castro, I.; Raman, A.; Sastry, N.; and Tyson, G. 2023. Flocking to Mastodon: Tracking the Great Twitter Migration. In *Proceedings of the 2023 ACM on Internet Measurement Conference*, IMC '23, 111–123. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703829.

Hoover, A. 2023. The Mastodon Bump Is Now a Slump. *Wired*.

Kiene, C.; Monroy-Hernández, A.; and Hill, B. M. 2016. Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1152–1156. New York, NY: ACM. ISBN 978-1-4503-3362-7.

King, J.-M. 2024. Mastodon Near Me.

Krasnoff, B. 2022. Mastodon 101: How to Follow (and Unfollow) Other Accounts. https://www.theverge.com/23519279/mastodon-instance-follow-friend.

Kraut, R. E.; Resnick, P.; and Kiesler, S. 2011. *Building Successful Online Communities: Evidence-Based Social Design*. Cambridge, Mass: MIT Press. ISBN 978-0-262-01657-5.

La Cava, L.; Aiello, L. M.; and Tagarelli, A. 2023. Drivers of Social Influence in the Twitter Migration to Mastodon. *Scientific Reports*, 13(1): 21626.

Nicholson, M. N.; Keegan, B. C.; and Fiesler, C. 2023. Mastodon Rules: Characterizing Formal Rules on Popular Mastodon Instances. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, 86–90. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701290.

Raman, A.; Joglekar, S.; Cristofaro, E. D.; Sastry, N.; and Tyson, G. 2019. Challenges in the Decentralised Web: The Mastodon Case. In *Proceedings of the Internet Measurement Conference*, IMC '19, 217–229. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6948-0.

Rochko, E. 2023a. Mastodon 4.2. https://blog.joinmastodon.org/2023/09/mastodon-4.2/.

Rochko, E. 2023b. A New Onboarding Experience on Mastodon. https://blog.joinmastodon.org/2023/05/a-new-onboarding-experience-on-mastodon/.

Roth, E. 2023. It's Getting Easier to Make an Account on Mastodon. https://www.theverge.com/2023/5/1/23707019/mastodon-account-creation-twitter-alternative.

Silberling, A. 2023. A Beginner's Guide to Mastodon, the Open Source Twitter Alternative | TechCrunch. https://techcrunch.com/2023/07/24/what-is-mastodon/.

TheKinrar. 2017. Mastodon Instances. https://instances.social/.

Trienes, J.; Cano, A. T.; and Hiemstra, D. 2018. Recommending Users: Whom to Follow on Federated Social Networks. arxiv:1811.09292.

Zulli, D.; Liu, M.; and Gehl, R. 2020. Rethinking the "Social" in "Social Media": Insights into Topology, Abstraction, and Scale on the Mastodon Social Network. *New Media & Society*, 22(7): 1188–1205.