

Exploring Platform Migration Patterns between Twitter and Mastodon: A User Behavior Study

Ujun Jeong¹, Paras Sheth¹, Anique Tahir¹, Faisal Alatawi¹, H. Russell Bernard², Huan Liu¹

¹School of Computing and Augmented Intelligence, Arizona State University

²Institute for Social Science Research, Arizona State University
{ujeong1, psheth5, artahir, faalataw, asuruss, huanliu}@asu.edu

Abstract

A recent surge of users migrating from Twitter to alternative platforms, such as Mastodon, raised questions regarding what migration patterns are, how different platforms impact user behaviors, and how migrated users settle in the migration process. In this study, we elaborate on how we investigate these questions by collecting data over 10,000 users who migrated from Twitter to Mastodon within the first ten weeks following the ownership change of Twitter. Our research is structured in three primary steps. First, we develop algorithms to extract and analyze migration patterns. Second, by leveraging behavioral analysis, we examine the distinct architectures of Twitter and Mastodon to learn how user behaviors correspond with the characteristics of each platform. Last, we determine how particular behavioral factors influence users to stay on Mastodon. We share our findings of user migration, insights, and lessons learned from the user behavior study.

Introduction

With the proliferation of social media platforms, users tend to be increasingly mobile, moving between different platforms as their needs, preferences, and interests change, a phenomenon known as online platform migration (Zengyan, Yinping, and Lim 2009). The ownership change of Twitter on October 27, 2022, created an opportunity to study platform migration as numerous users migrated to Mastodon (Kupferschmidt 2022; He et al. 2023), a microblogging platform with similar features to Twitter, such as “toots” and “boosts” corresponding to “tweets” and “retweets,” but which operates on a decentralized platform of self-hosted servers (Zignani, Gaito, and Rossi 2018).

Despite its decentralized architecture, Mastodon enables users to request follows and share toots, boosts, or favorite requests across servers using the shared Activity Pub protocol (Zignani, Gaito, and Rossi 2018). This facilitates the formation of federated servers and interest-based communities, each managed by distinct moderation policies, as depicted in Figure 1. Mastodon is thus not a replica of Twitter; it provides a unique, community-based user experience and interactions across various servers (La Cava, Greco, and Tagarelli 2021). Such distinctive features made Mastodon an increasingly popular choice for users seeking a new platform for migration.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

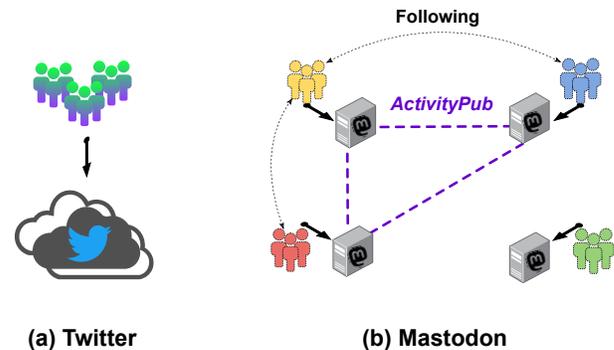


Figure 1: The distinct platform architectures of (a) Twitter, which is a centralized platform, and (b) Mastodon, which employs a decentralized platform with a federated network.

Online platform migration has been researched (Kumar, Zafarani, and Liu 2011; Newell et al. 2016; Fiesler and Dym 2020; He et al. 2023). Observing the exodus from Twitter, however, we are intrigued to understand user migration patterns: (1) what follows the exodus, i.e., whether migration is persistent or waning - this issue arises because users can maintain accounts on multiple platforms, oscillating their focus between platforms until fully committing to the new platform; (2) whether a platform architecture impacts the behaviors of migrated users - exploring this connection may reveal the interplay between platform differences, subsequent user behaviors post-migration, and varying degrees of user engagement; and (3) what behavioral factors contribute to the sustainability of platform migration is essential for a comprehensive grasp of the dynamics of the migration process.

In this study, we investigate migration patterns by comparing behaviors of users between Twitter and Mastodon, focusing on the dynamics of migrated users switching their attention between the two platforms over time. To facilitate this comparative study, we propose methods to map the accounts of over 10,000 individuals who migrated from Twitter to Mastodon and to determine users’ occupational backgrounds. Moreover, we collect data on user behaviors, including user activities and network interactions on each platform during the first ten weeks after Twitter’s leadership change.

Our study is motivated by three research questions:

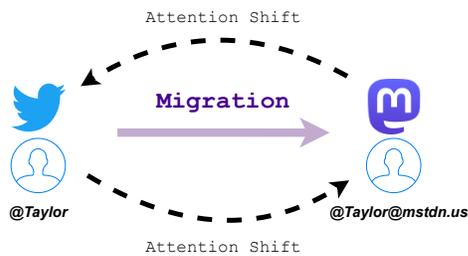


Figure 2: The process of migration, while still maintaining the previous account and shifting attention between platforms.

- **RQ1:** What are the migration patterns observed between Twitter and Mastodon after Twitter’s ownership change?
- **RQ2:** What distinct user behaviors between Twitter and Mastodon arise due to each platform’s architecture?
- **RQ3:** What specific behavioral characteristics of migrated users influence their decision to stay on Mastodon?

With respect to **RQ1**, we examine the evolving migration patterns with users’ active status and attention shifts between these platforms over time. Building on this analysis, we correlate migration patterns with notable events, such as Musk’s acquisition of Twitter, his promise of stepping down as a CEO, and so on. Through comparison of discourse and sentiment with migration patterns, we aim to uncover the potential motivations and how they affect migration process.

In light of **RQ2**, we compare the behaviors of migrated users on Twitter and Mastodon to understand the disparities between the two platforms. Especially, our experiments include comparing the distribution of users’ occupations and hashtags based on their popularity on each platform. This experiment aims to serve as empirical evidence of how the design of each platform affects the engagement of specific user groups more than others.

Regarding **RQ3**, we focus on the behavioral traits of users who stayed on Mastodon despite notable events drawing attention back to Twitter. By performing statistical analyses, we highlight associations between user retention and various user behaviors, including some distinctive features exclusive to Mastodon. Through this analysis, we shed light on Mastodon’s unique appeal to its users and draw insights on certain behavioral factors that promote sustainable migration.

Our main contributions are summarized as follows:

- Following the platform policies, we crawl and curate a dataset of over 10,000 users who migrated from Twitter to Mastodon by developing an effective method of mapping accounts between the two platforms.
- We propose a novel framework for understanding the intricate dynamics of migration patterns between Twitter and Mastodon, with a focus on the shift of user attention between the two platforms. It enables us to study the interplay among migration factors, the unique platform architectures, and the user behaviors on each platform.
- We present key behavioral factors that promote the sustainability of platform migration to Mastodon, interesting insights for future study of platform migration.

Related Work

Migration Theory and Platform Migration

Migration has long been a subject of study across social sciences, with the push-pull theory being a key concept in this field (Levitt and Jaworsky 2007; Lee 1966). The push-pull theory suggests that factors pushing people away from a location and pulling them toward a new one drive migration decisions. This concept can also be applied to online platform migration (Fiesler and Dym 2020; Gerhart and Koohikamali 2019; Newell et al. 2016). Newell et al. (2016) studied the motivations through push-pull factors by surveying users who migrated across subreddits during community unrest on Reddit, and found that niche communities play an essential role in attracting users. Factors including platform design, toxicity, moderation policy, presence of friends and community can be motivations for migration (Fiesler and Dym 2020). Through behavioral analysis, Kumar, Zafarani, and Liu (2011) found significant differences between migrants and random users in three key areas: posting activity, network activity, and Google search ranking. Lőrincz et al. (2017) found that social capital, such as a high degree of connection and openness of connection, can influence early abandonment of platform. He et al. (2023) studied the recent surge in users migrating from Twitter to Mastodon. This study presented several findings, including users’ tendency to join larger servers, the higher number of imported Twitter followers on these larger servers, and the fact that messages posted by migrated users had relatively lower toxicity on Mastodon than on Twitter.

Mastodon’s decentralized architecture attracted many users from traditional centralized platforms like Twitter and Facebook because of its several unique characteristics (He et al. 2023; La Cava, Greco, and Tagarelli 2021). La Cava, Greco, and Tagarelli (2021) found that connections in Mastodon are topic-oriented, rather than popularity-driven due to the lack of a recommendation system. Mastodon also emphasizes conversations and interactions over favorites and reshares (La Cava, Greco, and Tagarelli 2022b). Despite Mastodon’s unique appeals over traditional platforms, it faces numerous challenges inherent to decentralized networks. Information consumers on Mastodon establish most connections and broker information (La Cava, Greco, and Tagarelli 2022a). The user-driven trend toward the centralization of Mastodon, primarily on a handful of large servers, is also observed (Shaw 2020; Zignani, Gaito, and Rossi 2018). Other issues, such as advertising revenue, handling moderation, and the the availability of servers, pose challenges to operating Mastodon (Anaobi et al. 2023; Zulli, Liu, and Gehl 2020; Raman et al. 2019).

Our study diverged from prior platform migration research by examining dynamic migration patterns between Twitter and Mastodon over time. We focused on users’ attention shifts between platforms and presented algorithms to infer their migration motivations. Moreover, we analyzed the relation between platform architecture and the behaviors of migrated users, identifying behavioral factors that retain users on Mastodon. This holistic approach deepens our understanding of migration patterns and sustainable platform migration.

Migration Types on Social Media

In social media and migration studies, two types of migration are identified (Hou and Shiau 2020; Fiesler and Dym 2020; Kumar, Zafarani, and Liu 2011; Levitt and Jaworsky 2007): (1) *Permanent migration*, where users transition to a new platform, deactivate their original account, and exclusively engage on the new platform; and (2) *Attention migration*, where users maintain presence on both platforms but shift their focus toward one of them.

Permanent Migration. Let U_{p_1} represent the set of users on the platform p_1 and U_{p_2} represent the set of users on the platform p_2 . A user u is considered to have permanently migrated from platform p_1 to platform p_2 if two conditions are met: (1) the user u was a member of p_1 before time t , and (2) the user u is no longer a member of p_1 and is a member of p_2 at time t (i.e., $u \notin U_{p_1}$ and $u \in U_{p_2}$).

Permanent migration may be a result of profile removal, deletion, or suspension from the original platform.

Attention Migration. For a user u who is a member of both platforms (i.e., $u \in U_{p_1}$ and $u \in U_{p_2}$) and active on both platforms at time t_i , attention migration is said to occur between two distinct times t_i and t_j where $t_j > t_i$ if u becomes inactive on platform p_1 while remaining active on platform p_2 at time t_j . In this case, the user’s attention is considered to have migrated from platform p_1 to platform p_2 during the time interval $\delta = t_j - t_i$.

Attention migration could represent a short-term migration with the potential to evolve into permanent migration over time. To ascertain a user’s active status on a platform, we define an active user as:

Active User. Given a social media platform p , a user $u \in U_p$, time $t_j > t_i$, and interval $\delta = t_j - t_i$, u is considered to be active on the platform p at time t_j , if the user has performed at least one activity on the site since time t_i . Otherwise, the user is considered inactive. User activities include various actions possible on the social media platform, such as posting messages and resharing other users’ contents.

Our study recognizes that the permanent migration is not limited to the transition from Twitter to Mastodon, as the reverse direction is also possible. However, we focus on users who initiated migration from Twitter to Mastodon to examine the impact of Twitter’s ownership change on migration. For this specific objective, we excluded users who created Mastodon accounts before creating their Twitter accounts.

Data Collection

From October 24, 2022, to January 2, 2023, we identified a total of 10,333 migrated users who have accounts on both Twitter and Mastodon. We removed 56 users (0.5%) who had Mastodon accounts before Twitter accounts, along with 266 users (2.5%) whose accounts are inaccessible either on Twitter or Mastodon as of January 2, 2023. This process left us 10,011 migrated users for our analysis.

Mapping Users across Platforms

Several services, such as Twitodon, aid in merging Twitter and Mastodon accounts for users switching platforms but

Server Name	# User	# Migrant	Domain
mstdn.social	176,621	1,965	General
mastodon.world	146,753	1,245	General
mas.to	148,995	2,180	General
c.im	61,854	397	General
masto.ai	64,908	452	Technology
fosstodon.org	52,993	1,230	Technology
infosec.exchange	45,532	2,087	Technology
sfba.social	36,972	325	General
mindly.social	33,855	179	General
toot.community	28,227	273	General

Table 1: Top 10 Mastodon servers with most incoming users. We report the number of migrants detected by our method.

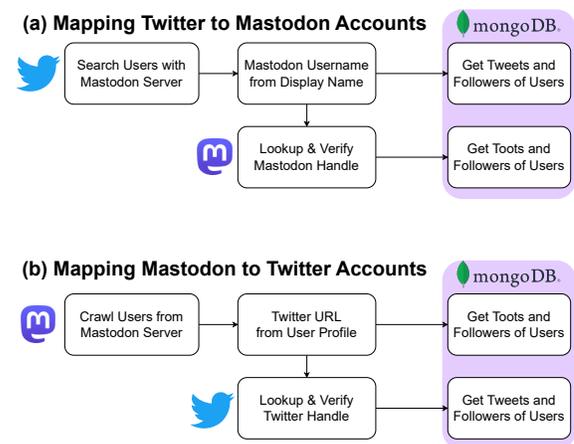


Figure 3: The process of data collection, including mapping accounts, using both Twitter and Mastodon as resources.

only support limited account mappings. He et al. (2023) addressed this by analyzing tweets containing Mastodon URLs, confirming matches only when Mastodon and Twitter usernames were identical. In contrast, our approach allows for the flexible identification of migrated users, even with nonidentical usernames, and incorporates both Mastodon and Twitter as resources for account mapping.

Mapping Twitter to Mastodon Accounts. We leverage Twitter users’ display names in their profile to map their Mastodon accounts, as these users often disclose their Mastodon usernames and associated Mastodon servers within their display names. We first search for users with the name of any Mastodon server listed in Table 1, and we extract the Mastodon usernames from the display names using a regular expression that identifies the word following the @ symbol.

Mapping Mastodon to Twitter Accounts. We utilize Mastodon users’ profiles, noting that they frequently disclose their additional social media accounts. We first collect user profiles from each Mastodon server. Then, we extract

Twitter handles from URLs in user profiles. If a Twitter URL is missing, we search the profile’s meta fields for entries marked with Twitter and extract the handle from this value.

It is worth noting that Mastodon’s decentralized structure poses challenges in gathering information across all servers (Raman et al. 2019). During our research, several prominent servers, such as *mastodon.social*, temporarily closed new account registrations due to a surge in user traffic. To maintain the integrity of this study, we concentrated on servers that are highly available and accessible when using the Mastodon API. Moreover, we utilized the *instances.social* API to analyze the number of users for each server. Specifically, we narrowed down our server selection to those primarily using English. This allowed us to gather data from the top 10 Mastodon servers that had the most newly registered users during the research period, as depicted in Table 1.

Collecting User Activities and Network Interactions

We utilized APIs, with their authenticated credentials, to monitor and collect various behaviors of the migrated users included in our study. The collected data were securely stored in a MongoDB database in an anonymized form, employing field-level encryption as a security measure.

Our data collection encompassed a variety of activities from detected migrants such as tweets, retweets, toots, and boosts. Our data collection efforts included gathering the messages posted during the research period on the public timelines of both Twitter and Mastodon. This process yielded 1.1M tweets and 0.5M retweets from Twitter, along with around 0.6M toots and 0.4M boosts from Mastodon.

We expanded our data collection to encompass followers of detected migrants on Twitter and Mastodon, resulting in obtaining nearly all available profiles of the followers of the migrants on these platforms: 85.42% of the total 5.3M followers on Twitter and 88.72% of the total 2.9M followers on Mastodon. We also collected data on interactions among the migrants on Mastodon, such as replies to the postings, and the profiles of users who participated in these replies.

Collecting and Grouping User Occupations

Since users may not always include occupations in their profiles, we utilized Stanford CoreNLP (Manning et al. 2014), a pretrained model capable of identifying occupation titles through named entity recognition. We first filtered users by examining Twitter profiles for occupation titles, and if these are unavailable, we referred to their corresponding Mastodon profiles. Then, we employed the UK’s Standard Occupational Classification (SOC 2010) system to assign an occupation per user, widely adopted in prior research on social media (Sloan et al. 2015). We assigned the user’s occupation title to the first corresponding three-digit SOC code from the system’s dictionary. If no match was found, we applied the method suggested by Turrell et al. (2019) utilizing the user’s profile bio. This method identified the SOC codes for 5,817 of 10,011 users (58.1%). To manage numerous SOC codes, we focused on the first digit (e.g., 1xx), which denotes the nine major groups. Two annotators classified these major groups for a random 300-user subset, resulting in a Cohen’s Kappa

of 77.82%. Our approach achieved F1 score of 65.26%, effectively predicting nine major groups. Figure 4 displays the distribution of occupational groups among migrated users, with 2xx (Professional Occupations) and 3xx (Associate Professionals and Technicians) groups being the most prominent.

RQ1: What Are the Migration Patterns between Twitter and Mastodon?

In this section, we investigate the extent to which individuals who were active on Twitter and Mastodon and how they shifted their attention to each platform following Musk’s acquisition of Twitter and other major events related to Twitter.

Understanding Evolving Migration Patterns

Figure 5 shows the trend between Twitter-related events and the number of daily active users, covering users active on one or both platforms within specified intervals. User activity is assessed over a one-day interval, starting at midnight of the preceding day (t_i) and ending at midnight of the current day (t_j). For example, setting t_j to January 2nd means the interval covers January 1st from 00:00 to 23:59.

To delineate major changes in Figure 5, we examine temporal shifts in daily active users using Prophet (Taylor and Letham 2018) developed by Facebook, a time-series method that is capable of handling nonlinear growth and seasonal changes. Our analysis involved applying Prophet separately to the three trends named Twitter-only, Mastodon-only, and Twitter & Mastodon. When pinpointing key moments where these temporal shifts overlapped, we found that these overlapped points correspond to Mastodon’s first peak (November 22, 2022), and its second peak (December 20, 2022). Based on these two important moments, we were able to distinguish three distinct phases in migration patterns: (1) the *protesting phase*, (2) the *adaptation phase*, and (3) the *current phase*.

The *protesting phase* began following the change in ownership of Twitter and concluded before the first peak in usage of Mastodon. During this phase, the number of active users on Twitter consistently declines, while there is a steep increase in the number of active users on Mastodon. These changes in the number of active users appear with various events happened on the Twitter, such as the mass layoffs at Twitter and the launch of Twitter Blue, which includes the subscription service for verified accounts and policy updates.

The *adaptation phase* follows the first peak and continues until the second peak of Mastodon. In response to recent policy changes on Twitter, there is a notable change in active user engagement on Mastodon. Especially, the number of active users on Mastodon temporarily surges after the relaunch of Twitter Blue and the suspension of Mastodon’s account and several high-profile journalists on Twitter¹. The similar upward trend is noticeable for users active on both platforms.

The *current phase* follows the second peak of Mastodon and continues until the present time. After Elon Musk announced his intention to step down as CEO of Twitter when he finds a suitable successor, the number of active users on

¹<https://www.nytimes.com/2022/12/18/business/twitter-ban-social-media-competitors-mastodon.html>

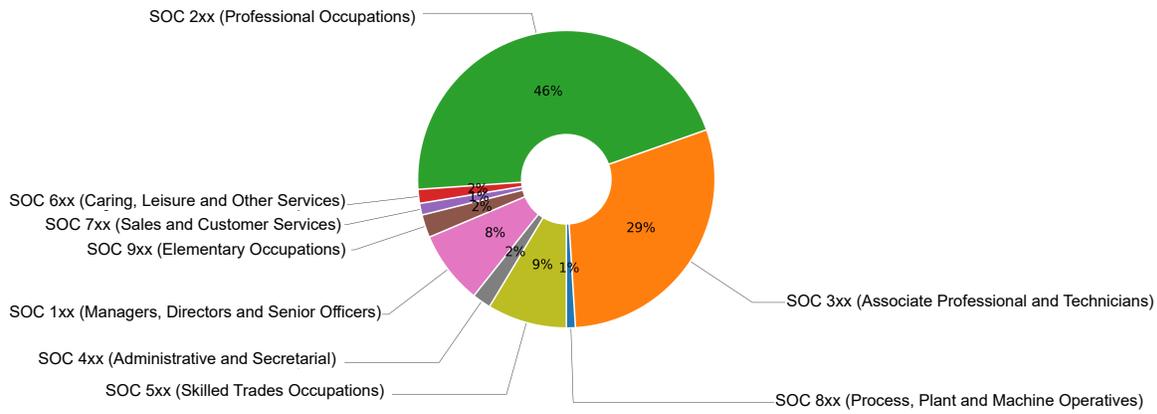


Figure 4: The pie chart illustrates the distribution of the distribution of nine major groups, each accompanied by corresponding tags. The major groups are based on the first digit of the UK's Standard Occupational Classification (SOC 2010) code, which has been assigned to each user. Further descriptions regarding nine major groups in the UK's SOC 2010 are provided in Appendix.

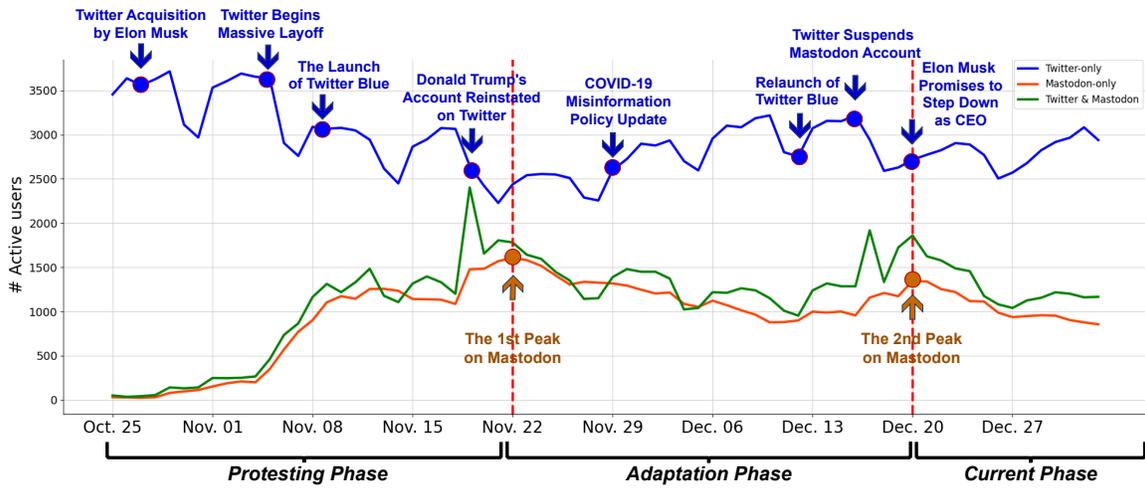


Figure 5: Trends in daily active users tagged with major events. Twitter-only (blue) and Mastodon-only (orange) lines indicate the number of users active exclusively on one platform, while Twitter & Mastodon (green) represents users active on both platforms. The x -axis denotes a particular date t_j , when we assess whether a user was active within the interval of $\delta = 1 \text{ day}$. Red dashed lines highlight the key moments, where the temporal shifts overlap among the three trends (blue, orange, and green).

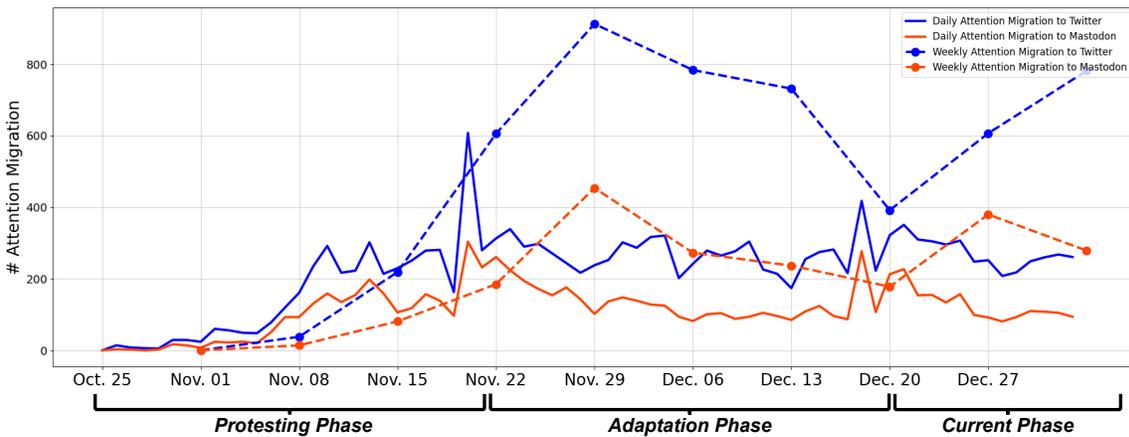


Figure 6: Trends in attention migrations toward Twitter (blue) and Mastodon (orange). The x -axis represents a date t_j when users migrated their attention. The trends are shown on a daily or weekly basis, with corresponding intervals of $\delta = 1 \text{ day}$ or $\delta = 1 \text{ week}$.

Mastodon started to decline. This event may have had a mitigating effect on attention given to Mastodon. At present, it appears that migrants tend to stay active on Twitter rather than continuing to use Mastodon, which raises questions about the longevity of the platform migration.

Figure 6 presents the trends of attention migration on daily and weekly intervals. These results indicate that the number of users shifting their attention to Twitter surpasses those moving to Mastodon, especially after each transition point between the migration phases. Notably, migrated users tend to shift more to Twitter after experiencing both platforms. The largest gap in attention migration trends between Twitter and Mastodon emerges during the *adaptation phase*. We speculate such a large gap in attention migration is because numerous users faced challenges in adjusting to Mastodon after arrival. Overall, the findings reveal that migrated users tend to stay on Twitter, and suggest that the momentum of migration to Mastodon has weakened.

Inferring Motivational Factors of Migration

To examine if any events on Twitter prompted users to migrate to Mastodon, we leveraged BERTopic (Grootendorst 2022) as a tool for examining temporal variations in their discourses. The frequency of tweets and toots relating to the top ten topic groups over time are graphically represented in Figure 7. Our experiment uncovered a significant portion of Twitter-related discourses (topic group 1 on both Twitter and Mastodon), as well as a strong focus on Elon Musk (topic group 4 on Twitter and topic group 1 on Mastodon). The prominent presence of Twitter-related discourses during the *protesting phase*, coupled with Elon Musk-related discourses toward the end of *adaptation phase*, offer quantitative evidence of a connection between Twitter’s change in ownership and the subsequent platform migration. Furthermore, our analysis revealed that 13.6% of the total toots included the keywords “Twitter” or “Elon Musk,” while only 6.1% of the total tweets contained these keywords. This difference shows that migrated users were proportionally more inclined to discuss Twitter and Elon Musk on Mastodon than on Twitter.

Identifying trending discourses, however, does not provide insight into the motivations for the migration. To fill this gap, we utilized sentiment analysis, leveraging a DeBERTa-based model (He et al. 2020; Yang et al. 2021), to investigate if Musk’s acquisition led users to leave Twitter as a form of protest. This model, adeptly fine-tuned for aspect-based sentiment analysis, allowed us to gauge sentiments specifically targeted at either Twitter or Elon Musk, in addition to capturing the general sentiment from a non-specific target.

Figure 8 illustrates the distribution of sentiment scores of tweets and toots on each platform. Our observations suggest that Mastodon may be particularly appealing to individuals with negative sentiments toward Twitter and its leadership. The migration from Twitter to Mastodon aligns with the push-pull theory of migration (Lee 1966), wherein the relatively negative sentiment toward Twitter and its leadership may act as a “push factor” (Newell et al. 2016) for users to leave Twitter. In contrast, users show more positive sentiment when Mastodon is a target than Twitter, which can serve as a “pull factor” attracting users to Mastodon (Hou and Shiau 2020).

RQ2: Distinct User Behaviors of Migrated Users on Twitter and Mastodon

Addressing this research question allows us to compare how migrated users exhibit their behaviors distinctively on Twitter and Mastodon, and discern if their varying behaviors can be explained by the unique platform architectures. Moreover, we also aim to uncover the disparities between these platforms based on the responses that users attracted on each platform.

User-level Features

To explore how users exhibit different behaviors on each platform, we decided to compare the common user behaviors that are shared between Mastodon and Twitter, such as tweets and toots. Given a specific user u and a platform p , we define three distinct types of user-level features on each platform: (1) *User Activity*, (2) *User Network*, and (3) *User Response*. The definitions for these categories are as follows:

User Activity. $\mathcal{A}(u, p)$ is determined by a user’s two activity types denoted as $\{a_i\}_{i=1}^2$, which comprise status messages and reshares (e.g., tweets and retweets a user has posted). The counting function $\sigma(u, a_i, p)$ measures the cumulative count of the activity a_i during the studied period. Formally, the user activity can be represented as:

$$\mathcal{A}(u, p) = \sum_{i=1}^2 \frac{\sigma(u, a_i, p) - \min_{u'} (\sigma(u', a_i, p))}{\max_{u'} (\sigma(u', a_i, p)) - \min_{u'} (\sigma(u', a_i, p))} \quad (1)$$

User Network. $\mathcal{W}(u, p)$ is derived by two network types denoted as $\{w_i\}_{i=1}^2$, which include followers and followees of a user. The counting function $\sigma(u, w_i, p)$ measures the size of the network w_i . The user network can be defined as:

$$\mathcal{W}(u, p) = \sum_{i=1}^2 \frac{\sigma(u, w_i, p) - \min_{u'} (\sigma(u', w_i, p))}{\max_{u'} (\sigma(u', w_i, p)) - \min_{u'} (\sigma(u', w_i, p))} \quad (2)$$

User Response. $\mathcal{R}(u, p)$ is obtained by two response types denoted as $\{r_i\}_{i=1}^2$, which include favorites and reshare a user has received (e.g., likes and retweets a user received). The counting function $\sigma(u, r_i, p)$ measures the cumulative count of the response r_i during the studied period. Formally, the user response can be represented as:

$$\mathcal{R}(u, p) = \sum_{i=1}^2 \frac{\sigma(u, r_i, p) - \min_{u'} (\sigma(u', r_i, p))}{\max_{u'} (\sigma(u', r_i, p)) - \min_{u'} (\sigma(u', r_i, p))} \quad (3)$$

where u' denotes a user within the population under study. Please note that we utilize max-min normalization on each platform when calculating $\mathcal{A}(u, p)$, $\mathcal{W}(u, p)$, and $\mathcal{R}(u, p)$. This normalization accounts for the different scales of user behaviors present on each platform.

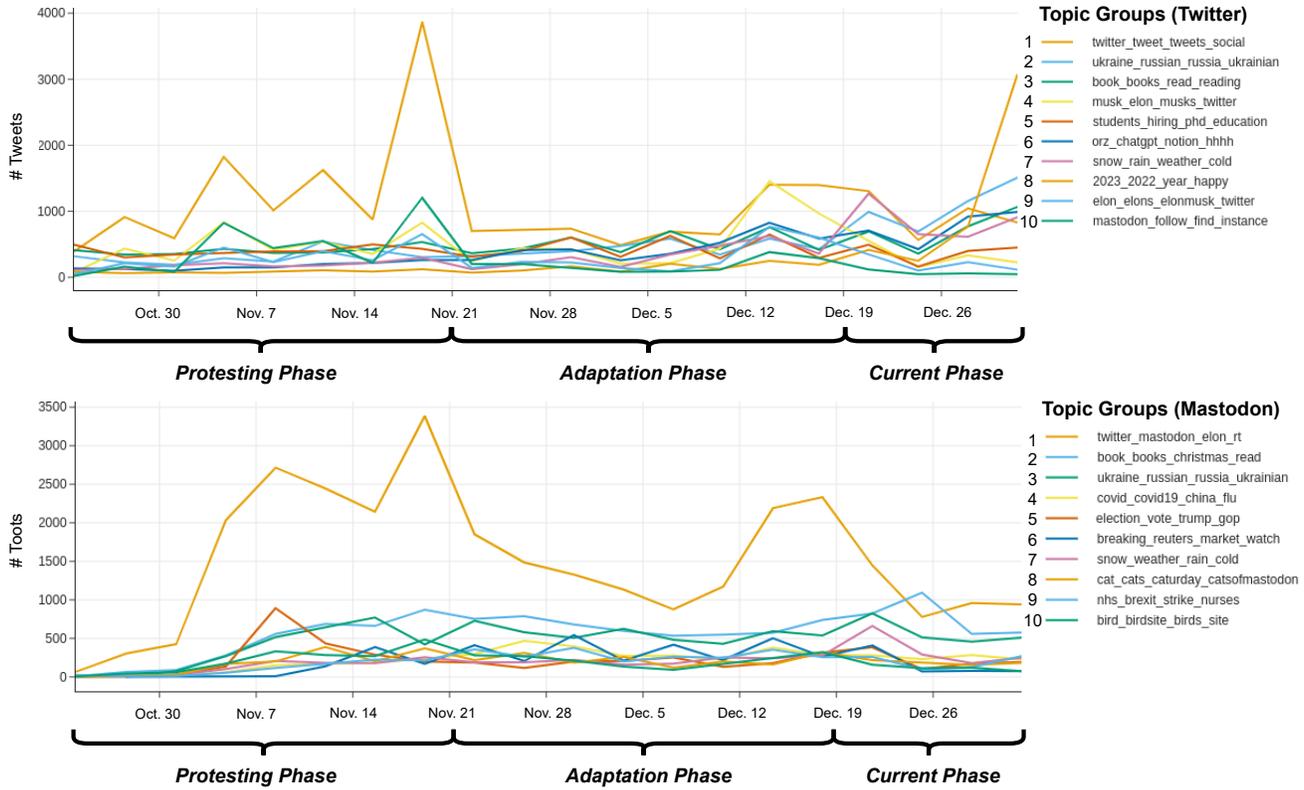


Figure 7: The trends in the number of tweets on Twitter and toots on Mastodon that mention one of the top 10 topic groups. Each topic group is shown with a discourse consisting of its four most frequently occurring words, separated by underscores. The topic groups are numbered based on their rank of proportion to the total messages, where a lower rank signifies a higher proportion. Note that the x -axis represents the exact date when the tweet or toot was published. As a result, the dates for the protesting, adaptation, and current phases are a day earlier than those referenced in Figure 5.

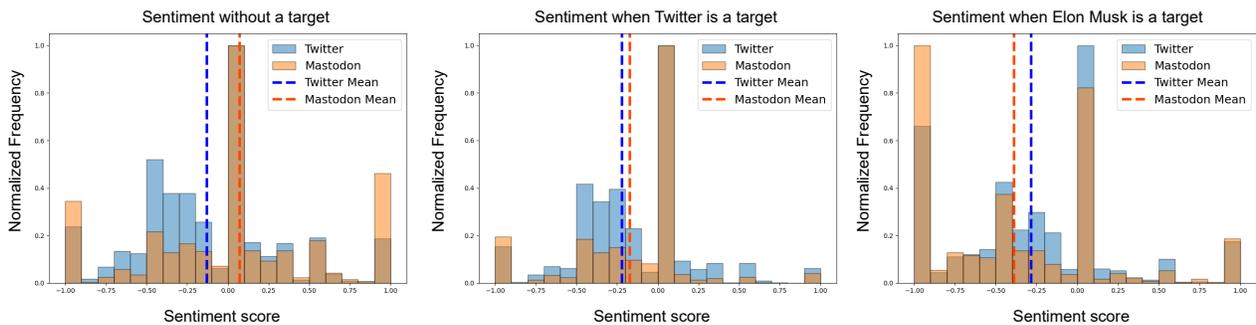


Figure 8: The sentiment score distribution is compared between Twitter (blue) and Mastodon (orange). We compare the distributions of Twitter and Mastodon across three categories: (1) sentiment without a target; (2) sentiment with Twitter as a target ; and (3) sentiment with Elon Musk as a target. The scores range from -1 to 1, with lower values indicating a more negative sentiment. The mean score for each distribution is shown by the dashed lines. The Kolmogorov-Smirnov test revealed significant sentiment differences between Twitter and Mastodon in all three categories, with a significant p -value of less than 0.001.

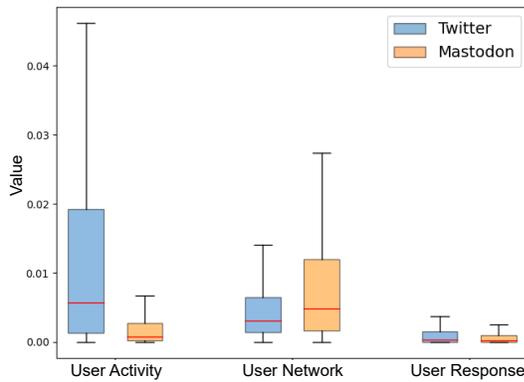


Figure 9: Interquartile box plots comparing the values of user-level features in Twitter (blue) and Mastodon (orange). Within each box plot, the red line indicates the median value.

Metric	\mathcal{A}		\mathcal{W}		\mathcal{R}	
	A	B	A	B	A	B
AVG	0.020	0.005	0.007	0.010	0.005	0.004
MAD	0.023	0.006	0.006	0.010	0.008	0.006

Table 2: Statistics of user-level features in platform **A** (Twitter) and **B** (Mastodon). AVG is average, and MAD (Mean Absolute Deviation) shows the dispersion from the average.

Disparity in User-level Features

We conducted a comparative analysis of three distinct user-level features on both Twitter and Mastodon platforms. Our comprehensive study derives insights from Table 2, which shows the mean and mean absolute deviation values, along with Figure 9, that illustrates the interquartile plot. Based on this analysis, we make three primary observations as follows:

First, we observe that users exhibit, on average, ~ 4 times higher *user activity* on Twitter than on Mastodon. This disparity may stem from the fact that numerous users continued to use both platforms even after migration, with Twitter serving as their primary platform (He et al. 2023). Additionally, the higher MAD value on Twitter indicates a broader spread of activities compared to Mastodon.

Second, we find that users show, on average, ~ 1.4 times more *user network* on Mastodon than on Twitter. This indicates that users tend to focus on building connections with other users on Mastodon (Shaw 2020) than Twitter. The MAD value is higher on Mastodon, signifying a more varied distribution of network sizes among the migrated users.

Third, our analysis reveal that users, on average, receive ~ 1.3 times more *user response* on Twitter than on Mastodon, likely due to Mastodon’s lower Q1 value (0.00003) compared to Twitter’s (0.003). The higher MAD value on Twitter suggests that user responses are more well spread among users on Twitter. Notably, a larger percentage of users receive zero responses on Mastodon (13.6%) than on Twitter (3.2%), consistent with previous findings of overall lower response rates on Mastodon (La Cava, Greco, and Tagarelli 2022b).

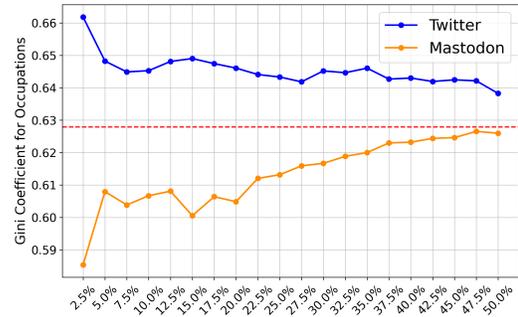


Figure 10: Gini coefficients for Twitter (blue) and Mastodon (orange) depict occupational distribution among migrants, focusing on users above the top 50% response rate. The red dashed line is the Gini coefficient for the original sample.

Disparity in Occupations to User Response

We utilized the Gini coefficient, a widely accepted metric for assessing disparities in income or wealth among diverse groups (Gini 1936), to evaluate the inequality in occupational distribution. The Gini coefficient values range between 0 and 1, with lower values signifying greater equality. We computed the Gini coefficient using n occupation groups (i.e., the nine major groups in the UK’s SOC code) and x_i , representing the user count for the i -th group among the 5,817 users (58.1% of the 10,011 users) with identifiable occupations as follows:

$$Gini\ Coefficient = 1 - \sum_{i=1}^n \left(\frac{x_i}{\sum_{j=1}^n x_j} \right)^2 \quad (4)$$

The Gini coefficient for 5,817 users is calculated to be 0.627. Such high inequality mainly stems from the overrepresentation of user groups like SOC 2xx (Professionals) and SOC 3xx (Associate Professionals and Technicians). One potential reason for this inequality is the unique appeal that Mastodon’s decentralized platform holds for individuals with academic or technology-related backgrounds (Kupferschmidt 2022; Shaw 2020; Zignani, Gaito, and Rossi 2018).

To discern if certain groups of occupations are encouraged to attract more user responses on each platform, we examined the occupational inequality of a subset in each platform. Our primary focus was on the relationship between occupational inequality and the user response on each platform. To this end, we first ranked users according to their responses on each platform. Once we created the ranking, we focused on the users who actively engage on each platform and these users represent a value above the median in the data.

Figure 10 demonstrates that occupational inequality on Twitter surpasses the inequality of the original sample within the top 50% of users. Interestingly, the disparity between Twitter and Mastodon peaks within the highest 2.5% of users, underlining a significant inequality among this extremely popular user group on Twitter. Conversely, on Mastodon, users exhibit lower inequality than the original sample up to the top 50% of users, suggesting a more evenly distributed occupational distribution among its highly engaged users.

Disparity in Hashtags to User Response

We compared the unique set of hashtags shared by each user with their popularity on Twitter and Mastodon. Given that a user’s popularity on one platform, gauged by the size of their user response, may not necessarily translate to popularity on another platform, we grouped users into four categories based on their popularity: (1) popular on Twitter, but not Mastodon; (2) popular on Mastodon, but not Twitter; (3) popular on both platforms; and (4) unpopular on both platforms.

Users were grouped into one of four categories based on their ranks on each platform, $rank_A$ for Twitter and $rank_B$ for Mastodon. For categories (1) and (2), users are sorted in descending order based on the difference in their ranks ($rank_{diff} = rank_A - rank_B$). This represents users with the largest differences in user response between the two platforms, with the top 10% being users who are popular on Mastodon but not on Twitter and the bottom 10% being users who are popular on Twitter but not on Mastodon. For categories (3) and (4), users are sorted based on the sum of their ranks ($rank_{sum} = rank_A + rank_B$). The top 10% in this case are users who are unpopular on both platforms, while the bottom 10% were users who are popular on both platforms.

Figure 11 showcases frequently used unique hashtags, classified into four categories based on the popularity of users on each platform. On Twitter, we observed popular users predominantly focus on social issue hashtags, such as #Covid and #Ukraine, whereas their less popular counterparts are inclined toward technology-oriented discussions, evident by those frequent hashtags such as #Opensource, #Security, and #Python. This suggests sharing global and social hashtags tend to be the main focus for getting more response on Twitter. In contrast, Mastodon users of all popularity levels engage with niche contents, encompassing photography, gaming, cats, and programming. Notably, users often share migration-specific hashtags on Mastodon, such as #TwitterMigration and #Introduction, which are relatively uncommon among the users popular on Twitter.

Analysis Summary

Our findings suggest empirical evidence indicating that the unique architectures of these platforms influence two aspects:

- The occupational inequality among popular users is more pronounced on Twitter than on Mastodon. This can be due to Twitter’s design, which focuses on popularity by promoting well-known users and their posts (La Cava, Greco, and Tagarelli 2021). In contrast, Mastodon’s topic-oriented approach creates connections and interactions based on shared interests. This approach provides an opportunity for migrated users from diverse occupations to gain popularity.
- Popular users on Twitter often share hashtags centered on global and societal issues. We speculate that this tendency could be linked to Twitter’s central structure, which enables broader reach and a larger audience, thereby amplifying the popularity of users with social significance (Kwak et al. 2010; Lehmann et al. 2012). Conversely, Mastodon users create connections by seeking out those with similar interests, promoting the sharing of more niche content, relatively irrespective of the users’ popularity levels.

Future research aims to enhance our understanding of the communication dynamics between Twitter and Mastodon. To investigate how information consumption is influenced among migrating users, further studies should focus on: (1) the impact of disparities between Twitter and Mastodon on the communication structure of both platforms; (2) the effects of disparities in occupational distribution on the overall user experience and information sharing; and (3) how these imbalances can inform strategies to improve user experience and promote equitable engagement among diverse users.

RQ3: Sustainability of Platform Migration: A Case Study of Mastodon

Behavioral Characteristics of Residents

In our investigation of user residency on Mastodon, we categorized the migrated users into two distinct groups: (1) *Non-residents*, whose last activity on Mastodon was either during the protesting or adaptation phase, and (2) *Residents*, whose last activity on Mastodon is during the current phase.

In analyzing the behavioral aspects that retain users on Mastodon, we applied user-level features to predict the two types of migrants. This investigation involved unique user-level features for Mastodon such as (1) Interaction Diversity, (2) Fandom Migration, and (3) Migration Hashtags.

Interaction Diversity. We conducted a study on interaction diversity for a specific user, denoted as u . This metric measures the level of interaction for a user u to engage with other users from different primary servers. The Shannon entropy formula is utilized to compute this metric as follows:

$$H(u) = - \sum_{i=1}^{|\mathcal{S}|} P(u, s_i) \log P(u, s_i), \quad (5)$$
$$P(u, s_i) = \frac{\sigma(u, s_i)}{\sum_{j=1}^{|\mathcal{S}|} \sigma(u, s_j)},$$

where $\sigma(u, s_i)$ is the cumulative count of interactions with users from server $s_i \in \mathcal{S}$, with interactions defined as receiving follows or replies by other users. $P(u, s_i)$ is the probability of user u interacting with users from server s_i .

Fandom Migration. We examined the concept of “Fandom migration” (Fiesler and Dym 2020), a phenomenon in which a leader’s migration to a new platform often leads to a collective migration of their followers from the previous platform as well. To determine the percentage of a user’s Mastodon followers who had previously followed them on Twitter, we analyzed identical usernames on both platforms, under the premise that identical usernames belong to the same person.

Migration Hashtags. To understand if users who led the migration from Twitter to Mastodon were more likely to remain, we analyzed the frequency of migration-specific hashtags, including: #RipTwitter, #GoodbyeTwitter, #JoinMastodon, #MastodonMigration, #MastodonSocial, #ByebyeTwitter, #TwitterMigration, #TwitterTakeover, #TwitterShutdown, #LeaveTwitter, #TwitterRefugee and #TwitterExodus.

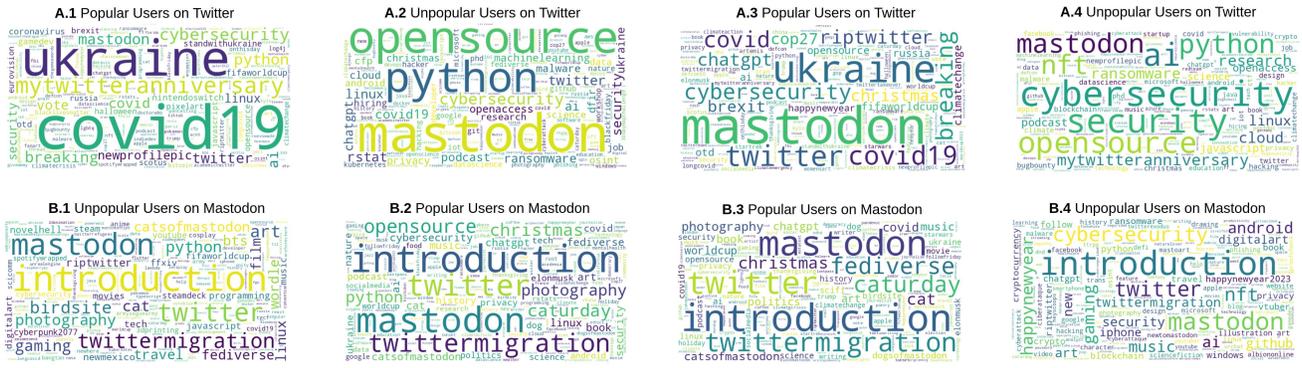


Figure 11: Wordclouds of unique lowercase hashtags on two platforms **A** (Twitter) and **B** (Mastodon). Hashtags are categorized based on the popularity of migrants: (1) popular on Twitter but not on Mastodon, (2) popular on Mastodon but not on Twitter, (3) popular on both platforms, and (4) unpopular on both platforms. More frequent hashtags are highlighted with larger font size.

For our analysis to classify the two types of migrants, namely *non-residents* and *residents*, we performed statistical tests on six distinct user-level characteristics. We found that a total of 5,590 *residents* (55.83% of the 10,011 users) were present at the time of the study. The outcomes of the logistic regression, after standardization of the feature values, are presented in Table 3. These results demonstrate the importance of various user-level features in predicting the types of migrants. Among the examined features, user activity, user network size, diversity of interactions, and fandom migration emerge as highly significant features, each demonstrating p -values lower than 0.001. Positive coefficients of these features reveal a positive association with the types of migrants, suggesting that as the value of these feature increases, so does the likelihood of migrated users being *residents*. Remarkably, user activity and interaction diversity show distinctively higher coefficients than other features, indicating their strong association with user retention to Mastodon.

Surprisingly, the volume of user response is not significantly associated with a user’s decision to stay on Mastodon, implying that users who merely focus on attracting a large user response tend to discontinue using the platform over time. Moreover, there is no statistical correlation between the frequency of sharing migration-specific hashtags and migrant types, indicating that individuals who once actively led the migration movement may not necessarily stay on Mastodon.

Analysis Summary

Mastodon, a decentralized social media platform, uses a federated network to encourage connections based on shared interests. However, this can lead to social fragmentation, complicating user interaction during building social networks. (Raman et al. 2019). Our analysis yields two notable outcomes:

- The decision to remain on Mastodon primarily is associated with a user’s active effort to frequent conversations and diverse social interactions to appreciate the community-centered experience, rather than by the volume of responses received or the frequent sharing of migration hashtags.
- Mastodon distinguishes itself by prioritizing community-centered experiences, contrasting with traditional platforms

Features	Coef.	SE	p -value	OR
User Activity	0.914	0.081	0.000**	2.494
User Network	0.173	0.040	0.000**	1.189
User Response	-0.053	0.032	0.100	0.948
Interaction Diversity	0.724	0.027	0.000**	1.077
Fandom Migration	0.074	0.022	0.000**	1.076
Migration Hashtags	-0.041	0.028	0.147	0.959

Table 3: Results on logistic regression to predict the two types of migrants (*non-resident* and *resident*) on Mastodon. Coef represents coefficient, SE stands for standard error, and OR denotes odd ratio. Note that $*p < 0.05$ and $**p < 0.001$.

that often give preference to individual self-promotion or the creation of viral content to capture a broader audience.

As a prospective avenue to enhance the sustainability of platform migration, it may be helpful to examine social network analysis theories that can shed light on concealed patterns within the social networks of migrated users by pinpointing information brokers linking disconnected groups, or structural holes (Burt 2018) and by examining the impact of weak ties in loosely connected networks on obtaining unique information and resources (Granovetter 1973). In turn, this information can inform platform developers, community managers, and users about the factors that facilitate sustainability of migration to a platform with distinct architecture and reduce mass departures and provide insight on how to avoid similar issues when launching new social media.

Limitations

First, the data available for this study do not include the 266 users who experienced *permanent migration* by January 2, 2023, due to API restrictions preventing access to deleted accounts on both Twitter and Mastodon. Our data is also limited to the first ten weeks post Elon Musk’s Twitter acquisition. As such, the findings from the migration patterns might not be applicable to other periods or predictive of future trends.

Second, we acknowledge the presence of alternative platforms for Twitter, such as Hive (a microblogging service) and Damus (decentralized network powered by Nostr protocol). However, our study focuses on the migration from Twitter to Mastodon, assuming that user attention mainly oscillates between Twitter and Mastodon due to Mastodon’s increased popularity during the time that Musk acquired Twitter.

Last, our study does not provide a causal analysis. Instead, the motivational factors are inferred from observed data on migration patterns, discourses over time, and users’ sentiment towards them. Testing the correctness of the inferred motivations for migration requires surveys of users including open-ended interviews, where users can report their reasons for migrating or not from any platform.

Conclusion

We introduce two types of migration on social media and explore the migration patterns between Twitter and Mastodon. Our observations indicate that dissatisfaction with Twitter and its management, notably the Twitter’s ownership change, is a primary motivation for users to migrate to Mastodon. Interestingly, these migrated users often utilize both platforms, frequently shifting their attention back to Twitter after trying both platforms. This migration pattern refutes claims of dooms day scenario for Twitter after the mass exodus.

Our comparison of user behaviors on Twitter and Mastodon unveils notable disparities between the two platforms. Specifically, users exhibit a larger inequality in the distribution of users’ occupations within highly engaged users on Twitter. An analysis of hashtag usage reveals that popular Twitter users typically concentrate on global and social issues, while Mastodon users, regardless of their popularity, often share hashtags related to migration and niche content.

We identify key user behaviors that are associated with the retention of users on Mastodon, with the variety of interactions with users across different Mastodon servers being a highly significant factor. Conversely, the volume of user responses, such as favorites and reshares, does not exhibit a significant impact on user retention. This underscores the distinctive appeal of Mastodon’s features, such as its decentralized architecture and community-oriented interactions, and how they relate to user behaviors.

In conclusion, our findings offer valuable insights for platform migration and how to retain users during periods after a mass migration. By leveraging these behavioral factors, platform designers can create more engaging and sustainable platforms that cater to user preferences and needs.

Future Work

Exploring the phenomenon of returning migration is an important area for future research. Our plan is to study this phenomenon as part of a long-term analysis when more user migration data becomes available between Twitter and Mastodon (Jeong et al. 2024). Furthermore, we aim to understand how the disparities between Twitter and Mastodon impact each platform, including user experience and the way information is shared among migrated users, including the spread of disinformation (Jeong et al. 2022a,b). One aspect

of this analysis is to apply theories of weak ties (Granovetter 1973) and structural holes (Burt 2018). Furthermore, surveys will be conducted to complement the findings and provide qualitative evidence of the potential motivations behind migration. Finally, the proposed approach will be extended to other alternative social media platforms beyond Twitter, to gather more comprehensive understanding of the migration.

Appendix

Standard Occupational Classification (SOC)

The UK’s SOC 2010 system uses a hierarchical structure to organize occupations, where the first digit of each SOC code represents the nine major groups, as shown in Table 4

Major Group	Description
1xx	Planing, directing, or coordinating the operations of businesses or other organizations.
2xx	Specialized tasks that require advanced knowledge, such as law, science, or education.
3xx	Support in various professional fields, such as healthcare, science, or engineering.
4xx	Administrative and clerical duties for business, government agencies or organizations.
5xx	Work in skilled jobs, such as construction, mechanics, or electrical work.
6xx	Personal services to individuals or groups, such as childcare, leisure, or personal care.
7xx	Selling products or services to customers or provide customer support.
8xx	Operating machinery or perform manual labor in manufacturing or production.
9xx	Simple and routine tasks that do not usually require formal education or training.

Table 4: Nine major groups in the UK’s SOC 2010 system.

Experimental Details

Utilizing the Prophet model set to change point prior scale of 0.252 and change point range of 0.95 with linear growth, we observed eight temporal shifts in Twitter-only users on specific dates (Nov. 1, 9, 22, 24, Dec. 5, 10, 20, 28), five in Mastodon-only users (Nov. 11, 19, 22, Dec. 7, 20), and five in Mastodon & Twitter users (Nov. 11, 19, 22, Dec. 5, 20). In BERTopic model, we utilized `all-MiniLM-L6-v2` as the default model, which is a pre-trained transformer model designed for embedding English sentences and short paragraphs. Focusing on toots and tweets in English, we optimized our system settings to auto-adjust with a minimum of 100 topics, which resulted in generating 841 topics for tweets and 483 for toots. For aspect-based sentiment analysis, we leveraged the DeBERTa with `deberta-v3-large-absa`, which is a fine-tuned model for ABSA (Aspect-Based Sentiment Analysis) datasets. The model’s default parameter setup has a hidden size of 768, 12 hidden layers, and 12 attention heads.

Ethical Statement, Impact, and Reproducibility

Our proposed approach for analyzing migration from Twitter to Mastodon may suffer from selection bias, as it relies on

users who have chosen to disclose their other social media platform accounts on their profiles. As a result, the selected users may not be representative of the larger population of users migrating from Twitter to Mastodon.

Our study aims to gain insights into the migration of users between Twitter and Mastodon, and is not intended to exploit or manipulate these communities for financial gain. Our data collection was completed before Twitter updated its basic API tier to a priced model. In compliance with the terms of service of Twitter and Mastodon, we disclose only the user IDs of the accounts included in our study. The codes used for mapping user accounts between Twitter and Mastodon are available at our repository²

Acknowledgments

This work received support from the Office of Naval Research (Award No. N00014-21-1-4002) and the National Science Foundation (Award No. 2227488). Opinions, interpretations, conclusions, and recommendations within this article are solely those of the authors. Special acknowledgment goes to Dr. Kaize Ding, Assistant Professor at Northwestern University, whose insightful feedback was instrumental in improving the quality of the article.

References

- Anaobi, I. H.; Raman, A.; Castro, I.; Zia, H. B.; Ibosiola, D.; and Tyson, G. 2023. Will Admins Cope? Decentralized Moderation in the Fediverse. In *WWW*. ACM.
- Burt, R. S. 2018. Structural holes. In *Social Stratification*. Routledge.
- Fiesler, C.; and Dym, B. 2020. Moving across lands: Online platform migration in fandom communities. *Human-Computer Interaction*.
- Gerhart, N.; and Koohikamali, M. 2019. Social network migration and anonymity expectations: What anonymous social network apps offer. *Computers in Human Behavior*.
- Gini, C. 1936. On the measure of concentration with special reference to income and statistics. *Colorado College Publication*.
- Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology*.
- Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- He, J.; Zia, H. B.; Castro, I.; Raman, A.; Sastry, N.; and Tyson, G. 2023. Flocking to Mastodon: Tracking the Great Twitter Migration. In *Internet Measurement Conference*. ACM.
- He, P.; Liu, X.; Gao, J.; and Chen, W. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.
- Hou, A. C.; and Shiau, W.-L. 2020. Understanding Facebook to Instagram migration: a push-pull migration model perspective. *Information Technology & People*.
- Jeong, U.; Alghamdi, Z.; Ding, K.; Cheng, L.; Li, B.; and Liu, H. 2022a. Classifying COVID-19 related Meta ads using discourse representation through a hypergraph. In *SBP-BRiMS*. Springer.
- Jeong, U.; Ding, K.; Cheng, L.; Guo, R.; Shu, K.; and Liu, H. 2022b. Nothing stands alone: Relational fake news detection with hypergraph neural networks. In *IEEE Big Data*. IEEE.
- Jeong, U.; Nirmal, A.; Jha, K.; Tang, X.; Bernard, H. R.; and Liu, H. 2024. User Migration across Multiple Social Media Platforms. In *SDM*. SIAM.
- Kumar, S.; Zafarani, R.; and Liu, H. 2011. Understanding user migration patterns in social media. In *AAAI*. AAAI Press.
- Kupferschmidt, K. 2022. As Musk reshapes Twitter, academics ponder taking flight. *Science*.
- Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *WWW*. ACM.
- La Cava, L.; Greco, S.; and Tagarelli, A. 2021. Understanding the growth of the Fediverse through the lens of Mastodon. *Applied network science*.
- La Cava, L.; Greco, S.; and Tagarelli, A. 2022a. Information consumption and boundary spanning in decentralized online social networks: the case of mastodon users. *Online Social Networks and Media*.
- La Cava, L.; Greco, S.; and Tagarelli, A. 2022b. Network analysis of the information consumption-production dichotomy in mastodon user behaviors. In *ICWSM*. AAAI Press.
- Lee, E. S. 1966. A theory of migration. *Demography*, 3: 47–57.
- Lehmann, J.; Gonçalves, B.; Ramasco, J. J.; and Cattuto, C. 2012. Dynamical classes of collective attention in twitter. In *WWW*. ACM.
- Levitt, P.; and Jaworsky, B. N. 2007. Transnational migration studies: Past developments and future trends. *Annu. Rev. Sociol.*
- Lőrincz, L.; Koltai, J.; Győr, A. F.; and Takács, K. 2017. Collapse of an online social network: The blame on social capital. Technical report, IEHAS Discussion Papers.
- Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J. R.; Bethard, S.; and McClosky, D. 2014. The Stanford CoreNLP natural language processing toolkit. In *ACL system demonstrations*.
- Newell, E.; Jurgens, D.; Saleem, H.; Vala, H.; Sassine, J.; Armstrong, C.; and Ruths, D. 2016. User migration in online social networks: A case study on reddit during a period of community unrest. In *ICWSM*. AAAI Press.
- Raman, A.; Joglekar, S.; Cristofaro, E. D.; Sastry, N.; and Tyson, G. 2019. Challenges in the decentralised web: The mastodon case. In *Internet Measurement Conference*. ACM.
- Shaw, C. R. 2020. Decentralized Social Networks: Pros and Cons of the Mastodon Platform.
- Sloan, L.; Morgan, J.; Burnap, P.; and Williams, M. 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS one*.
- Taylor, S. J.; and Letham, B. 2018. Forecasting at scale. *The American Statistician*.
- Turrell, A.; Speigner, B. J.; Djumalieva, J.; Copple, D.; and Thurgood, J. 2019. Transforming naturally occurring text data into economic statistics: The case of online job vacancy postings. Technical report, National Bureau of Economic Research.
- Yang, H.; Zeng, B.; Xu, M.; and Wang, T. 2021. Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning. *arXiv preprint arXiv:2110.08604*.
- Zengyan, C.; Yinping, Y.; and Lim, J. 2009. Cyber migration: An empirical investigation on factors that affect users' switch intentions in social networking sites. In *System Sciences*. IEEE.
- Zignani, M.; Gaito, S.; and Rossi, G. P. 2018. Follow the "mastodon": Structure and evolution of a decentralized online social network. In *ICWSM*. AAAI Press.
- 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*.

²<https://github.com/ujeong1/ICWSM24-Exploring-Platform-Migration-Patterns-between-Twitter-and-Mastodon>

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes, the abstract provides three research questions, and the questions are mentioned and solved in each section respectively.](#)
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? [Yes, justification is provided for each methodology when answering research questions.](#)
 - (d) Do you clarify what possible artifacts in the data used, given population-specific distributions? [Yes, see the Limitations.](#)
 - (e) Did you describe the limitations of your work? [Yes, see the Limitations.](#)
 - (f) Did you discuss any potential negative societal impacts of your work? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (g) Did you discuss any potential misuse of your work? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? [NA](#)
 - (b) Have you provided justifications for all theoretical results? [NA](#)
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? [NA](#)
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? [NA](#)
 - (e) Did you address potential biases or limitations in your theoretical framework? [NA](#)
 - (f) Have you related your theoretical results to the existing literature in social science? [NA](#)
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? [NA](#)
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? [NA](#)
 - (b) Did you include complete proofs of all theoretical results? [NA](#)
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [NA](#)
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [NA](#)
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [NA](#)
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? [Yes, see the Data Collection.](#)
 - (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? [NA](#)
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
 - (a) If your work uses existing assets, did you cite the creators? [Yes, see the Data Collection and Research Questions \(RQ1 and RQ2\) for the assets we used for the research.](#)
 - (b) Did you mention the license of the assets? [No, but we confirmed BERTopic, DeBERTa, StanfordCoreNLP, and Prophet are available under the MIT License. The occupation encoder used in this paper utilizes the UK's SOC system and the code is under the GNU General Public License.](#)
 - (c) Did you include any new assets in the supplemental material or as a URL? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [NA](#)
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes, see the Ethical Statement, Impact, and Reproducibility.](#)
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? [No, because Twitter has changed its API to a pricing model, it limits accessibility to the protocol at the moment.](#)
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? [No, it is not available at this time due to the terms and conditions on Twitter \(or X\) for developers. See the Ethical Statement, Impact, and Reproducibility sections for more details.](#)
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
 - (a) Did you include the full text of instructions given to participants and screenshots? [NA](#)
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? [NA](#)
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [NA](#)
 - (d) Did you discuss how data is stored, shared, and deidentified? [NA](#)