Debirdifycation and the rise of Mastodon: a Computational Social Science analysis of the Twitter exodus.

Stefano Pardini*
University of Trento
stefano.pardini@studenti.unitn.it

1 INTRODUCTION

In the dynamically evolving landscape of social media, Decentralized Online Social Networks (DOSNs), with Mastodon being one of the most significant representatives, are emerging as compelling alternatives to traditional centralized platforms. This shift is not occurring in isolation but is part of a broader transformation in the social media domain where platforms like Twitter play a significant role.

This transition has been energized by key events, with Elon Musk's acquisition of Twitter serving as a noteworthy example. This specific event led to a significant surge in user registrations on Mastodon, hinting at potential migration patterns and their implications for the broader digital social landscape.

The remainder of the study maintains its focus on the notable increase in Mastodon user registrations following these major social media events, offering insights into potential migration trends and their implications for the digital social landscape. We investigate the correlation between these events, examining consequential user behaviors and themes associated with this platform migration.

1.1 Mastodon: A Case Study in DOSN

Mastodon, one of the most significant representatives of DOSNs, is steadily becoming a popular substitute for mainstream centralized platforms. The integral pillars of DOSNs, including open-source software and specialized communication protocols fostering seamless inter-server connections, culminate in a federated model where different servers, referred to as *instances*, can correspond with each other. This federated structure forms a vast network known as the *Fediverse*. Owing to this structure, users can interact across various platforms using their Mastodon account, eliminating the need for additional accounts.

Unlike proprietary services, "anyone has the complete freedom to run, examine, inspect, copy, modify, distribute, and reuse the Mastodon source code, provided they guarantee the same freedoms for any derivative work" [10].

2 LITERATURE REVIEW

The last decade has seen DOSNs, particularly Mastodon, carving their niche in the digital space by providing an alternative to mainstream centralized platforms. Mastodon's federated model, which fosters seamless interaction across different servers, culminates in a sprawling network known as the Fediverse. These user-centric features that prioritize conversation over popularity have been attractive to users migrating from platforms like Twitter and Facebook [8, 19].

Several studies have dedicated efforts to examine user migration patterns from centralized platforms to DOSNs. These migrations are influenced by factors ranging from data privacy concerns and dissatisfaction with community policies to significant events like changes in platform ownership, such as Musk's acquisition of Twitter [16, 21]. A variety of methodologies have been employed to scrutinize these migrations, such as content analysis of posts and examination of user engagement and behavior on the new platform [3, 17].

Yet, the decentralized model presents its own set of challenges such as centralization trends, advertising revenue obstacles, and moderation tasks [1, 12, 13, 20, 22]. A comprehensive understanding of these dynamics and the reasons for user migration is necessary for comprehending the rapidly evolving digital landscape [7].

3 RESEARCH QUESTIONS

In order to capture a comprehensive understanding, we've formulated three research questions which jointly examine the correlations between key events and user registrations, engagement levels of new users, and the primary themes and patterns driving user migrations.

3.1 RQ1

How do key events related to Elon Musk's acquisition of Twitter correlate with surge in user registrations on the alternative platform Mastodon, potentially signifying user migration trends?

This research aims to explore the relationship between pivotal developments in Elon Musk's takeover of Twitter and the subsequent increase in user registrations on Mastodon. The core element at the heart of this investigation is to understand whether actions related to the Twitter acquisition, such as Musk's public announcements or significant company changes, might act as catalysts, triggering a migratory movement of users to Mastodon.

By exploring a temporal window that encompasses both the period before and after the official acquisition date of October 27, 2022, the investigations aim to offer a comprehensive perspective on Musk's high language performativity. Can a single *flutter of Musk's wings* generate significant and measurable correlated effects, not only in financial terms?

3.2 RQ2

How does the activity of newcomers, as evidenced by their engagement in public toots within the most representative Mastodon federated instance, correlate with the daily enrollment rate of new users on the platform?

This second research question extends our exploration of user behavior on Mastodon, focusing specifically on the engagement levels of newcomers in cohorts based on their registration date. Engagement is gauged by the participation of these new users in public toots within Mastodon's

*All codes related to this research paper are publicly available at https://github.com/s-pardox/CSS-project.

most populated federated instance, *mastodon.social*. By correlating this engagement with the daily rate of new user registrations, our analysis aims to discern whether an increase in user registrations is associated with a proportionate rise in platform activity.

3.3 RQ3

What are the primary themes and patterns evident in the reasons users provide for migrating from Twitter to Mastodon?

This question guides an exploration of user discourse, underpinned by a content analysis of tweets and 'toots' tagged with *#mastodonmigration* and *#twittermigration*.

The aim is to identify the prevalent themes in users' self-disclosed motivations for transitioning between platforms and discern the patterns inherent in these public disclosures.

4 RESEARCH METHODOLOGY

4.1 Data collection

Step 1: Users statistics. Our first step in the data collection process involved obtaining statistics on the number of new users registering on Mastodon, focusing primarily on the time frame of the mass exodus. This data offered insights into the magnitude of the migration and allowed us to pinpoint when surges in new user registrations occurred.

The user registration data was gathered using two distinct methods and different periods of interest:

1. **Toot retrieval and parsing:** We obtained global user statistics by interacting with the Mastodon API to retrieve and parse public toots from the accounts *mastodonusercount@bitcoinhackers.org* and subsequently *mastodonusercount@mastodon.social*. The period of data collection spanned from February 1, 2022, to May 15, 2023. Both of these accounts provide hourly public toots containing user registration statistics, which allowed us to track and analyze the overall user registration trends on Mastodon over the defined period.
2. **Instance information extraction:** Additionally, we retrieved the `instances.json` file from the repository *simonw/scrape-instances-social* by querying the GitHub API. This repository, through an automated background process, periodically downloads the central `instance.json` file approximately every 30 minutes. The JSON file contains detailed information about users across different Mastodon instances, offering us granular data on user registration at the instance level. For the purposes of our research, we ingest data from the earliest public commit (November 20, 2022) up to May 15, 2023.

These combined approaches, spanning overlapping time periods, allowed us to form a comprehensive understanding of the user registration landscape during the period of interest, providing the foundation for our subsequent analysis.

Step 2: Hashtag related toots ingestion. The second major step in our data retrieval process was interacting with the Mastodon API to pull all the toots from the federated timeline related to two specific hashtags: *#mastodonmigration* and *#twittermigration*. These hashtags were chosen as they were deemed most likely to be representative of the migration discourse. The time span for this retrieval was from the inception of these hashtags until May 15, 2023.

During this data retrieval process, we were mindful of potential limitations and biases that might affect the representativeness of our dataset. Not all users who migrated to Mastodon may have used these hashtags in their toots, and the use of these hashtags might vary across different instances and user groups. Therefore, while these toots offered a rich and valuable source of data, we acknowledged that they might not fully capture the complete range of experiences and perspectives during the migration period.

The retrieved data was initially saved in JSON format on disk for further analysis and processing.

Step 3: Public toots retrieval. The third major step in our data collection process involved retrieving all public toots from the *mastodon.social* public timeline. The span of this retrieval covered a one-month period, from November 20, 2022, to December 19, 2022. This specific period was chosen because it coincided with the highest user registration peaks and the main events surrounding Elon Musk's Twitter acquisition observed within the Mastodon network.

This phase posed a unique challenge due to the substantial volume of data and the need to comply with the Mastodon API's request limitations.

Given the volume of the data, we designed and deployed a data processing pipeline for efficient ingestion. This pipeline interacted with the Mastodon API, captured the data in JSON format and stored it on disk, then processed the data and saved it in MySQL tables. The raw data derived from this process was approximately 13GB in size.

This process presented us with a comprehensive snapshot of the public activity on *mastodon.social* during the period of interest, offering insights into the broader ecosystem beyond the migration-related discourse captured through the *#mastodonmigration* and *#twittermigration* hashtags. However, we were aware that this data only represented a portion of the total activity on Mastodon, as it didn't capture private toots or interactions occurring in specific instances.

The selection of *mastodon.social* was guided by a combination of strategic and practical considerations. Strategically, this instance is the most populated within the Mastodon network, as shown in Figure 1, and experienced the most significant influx of new user registrations during the Twitter exodus, making it a compelling hub for studying post-migration user behaviors and patterns.

4.2 Text preprocessing

Influenced by Symeonidis et al.[14] and Ramachandran and Parvathi[11], we streamlined the toot data. After loading it into a DataFrame, we conducted manipulations such as parsing usernames, unescaping HTML, and eliminating certain entities, sequences, and non-ASCII characters. The content was then formatted for analysis, including language assignment (retaining only English toots), stopword removal, and TF-IDF filtering to distill high-impact terms [14]. We discarded toots with minimal content, added

https://github.com/simonw/scrape-instances-social
https://instances.social/

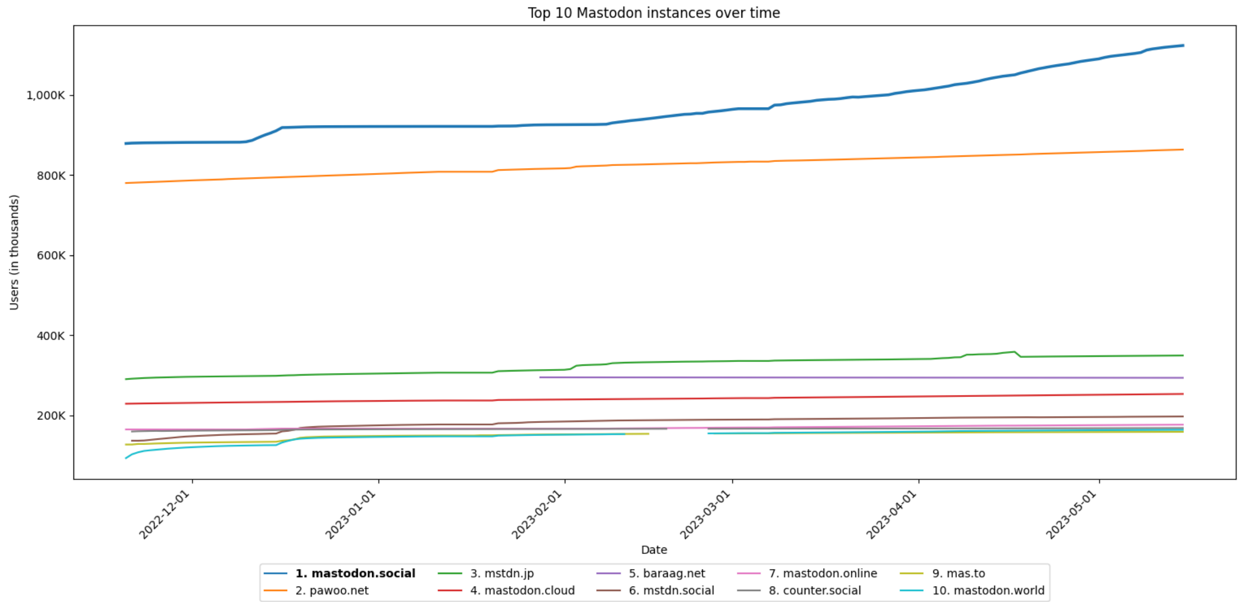

Figure 1: Top 10 Mastodon instances over time

a *date* field for referencing, and saved the result as a ready-to-analyze CSV file.

4.3 Data analysis

Data analysis was conducted in several phases, corresponding to the research questions presented.

Throughout our data analysis, we adhered to ethical guidelines to ensure that our research was conducted responsibly and with respect for the privacy of Mastodon users. In particular, we ensured that any data used in this study was publicly available and did not include any personally identifiable information. Additionally, all findings are presented in aggregate form to prevent the identification of individual users.

Phase 1: Analysis of user registration data. In investigating the first research questions, we consider the ongoing impact of Musk's decisions throughout the entire process of Twitter's acquisition. The overarching goal is to elucidate the ways in which corporate-level decisions, especially those with high visibility and potentially controversial, can influence user behavior within the digital domain either directly or indirectly. By employing qualitative Web-based research methods, and analyzing Tim Chambers' reports [5], [6], and [4], we have chosen the most pivotal events related to Musk's engagement with Twitter (Table 1). We hypothesize that these events, which occurred one or a few days prior, significantly influenced the volume of new registrations on the Mastodon fediverse, leading to noticeable spikes as shown in Figure 2. These events could potentially drive user migration trends.

At the crossroads of social media dynamics and high-stakes corporate maneuvering, the statements and actions of influential individuals assume a performative dimension, transforming mere words into consequential deeds [9].

Examining the results in this context provides a compelling perspective, illuminating the correlation between Musk's announcements and shifts in social media usage, while also spotlighting the socio-economic power tied to Musk's persona. This power, arising from his status as the world's wealthiest individual, amplifies the impact of his words and actions beyond mere statements. Thus, his performative power could prompt substantial shifts in user behaviors, potentially triggering a migration from one social media platform to another.

We are aware that for a more thorough and objective analysis, it would be beneficial to incorporate data from Twitter users in addition to the Mastodon data already gathered. While our current analysis of the Mastodon data allows us to infer a potential migration of users, a more precise understanding of this phenomenon would require an analysis of accounts that were actually closed on Twitter.

Phase 2: Analysis of newcomer engagement. Our findings contribute to understanding user behavior on Mastodon, focusing on the engagement levels of newcomers. The analysis was specifically carried out on the most populous federated instance, *mastodon.social*, given its representative nature in the Mastodon ecosystem. Using the retrieved raw data, we specifically elaborated and analyzed the following data fields:

- **enrollment_day:** Account creation date.
- **incremental_new_users:** Day-to-day net increase in new users.
- **toots_on_enrollment_day:** Unique users who posted on their account creation day.
- **toots_in_next_3_days:** Unique users who posted within three days post-enrollment, excluding enrollment day.

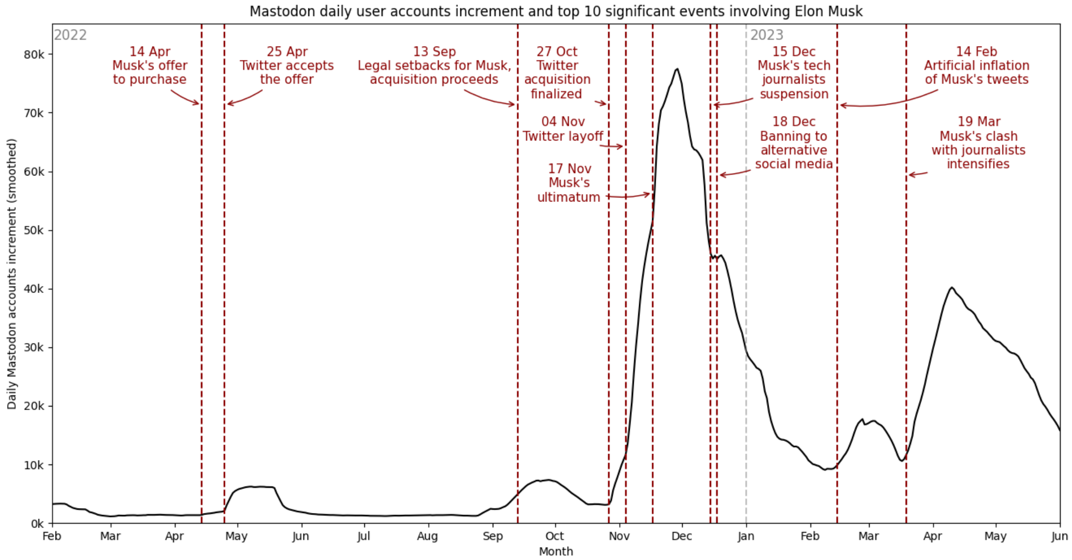

Figure 2: Top 10 significant events involving Elon Musk that correlate with spikes in user registration on Mastodon.

- **toots_days_4_to_7:** Unique users who posted between the 4th and 7th days post-enrollment, excluding previous days.

In the analysis of user enrollment and engagement patterns on the Mastodon platform, we have found several key trends. Firstly, the day-to-day growth of new users on the platform shows a very strong positive correlation with the number of distinct new users who made toots on the same day of their enrollment, with a correlation coefficient of approximately 0.986. This indicates a significant tendency for new users to start participating actively in the platform on the same day as their enrollment.

Furthermore, our analysis reveals an average engagement ratio (the ratio of toots_on_enrollment_day to incremental_new_users) of approximately 0.211. This suggests that around 21% of the new users on any given day engage in making toots on the same day of their enrollment.

Interestingly, as we expand the timeframe to include user activity for several days post-enrollment, we notice a reduction in the correlation with *incremental_new_users*. The correlation is approximately 0.980 for user tooting activity in the second, third days following enrollment, and drops slightly to approximately 0.958 for activity in the fourth to seventh days.

It's important to remember, however, that this analysis makes an implicit assumption. Each time we record activity within a subsequent time window (like *toots_in_next_3_days* and *toots_days_4_to_7*), we're continuing to register the activity of potentially the same users. This is a form of aggregation and therefore, it introduces an approximation into our analysis. Thus, while the overall trend of decreasing engagement over time is informative, this aspect of the methodology should be kept in mind when interpreting the results.

Moreover, it is worth mentioning that the present analysis measures engagement by observing the published toots. It does not account for forms of *passive participation* such as logging in, reading toots, following other users, sending friend requests, etc. These actions can also indicate a user's engagement with the platform but are not captured in the current dataset and analysis.

In summary, our findings indicate a robust level of engagement among Mastodon users. Despite the decrease in activity over time, the initial interaction remains strong, underscoring Mastodon's capacity to engage new users effectively.

Phase 3: Thematic analysis of migration-related discourse. To uncover latent topics within the toots, we utilized the Latent Dirichlet Allocation (LDA) and hashtag frequency analysis. LDA, a widely used machine learning technique for topic modeling [2], proved challenging due to the complexity of extracting semantic themes from the identified keyword collections. The details of these themes and the keyword collections are outlined in Table 2.

We focused our analysis on the period from 2022/10/20 to 2023/01/01, a time-frame marked by significant events that, as we can see in Figure 3, are highly correlated with a rise in the use of hashtags *#mastodonmigration* and *#twittermigration*.

We employed LDA within a narrow three-day window post-events, aiming to uncover event-related topics and future implications. Yet, even with this strategic focus, the topics' semantic complexity posed a significant interpretative challenge.

This highlights the difficulty of automatic topic extraction from social media discourse, demonstrating that despite LDA's success, its results can be obscure when applied to nuanced social media conversations.

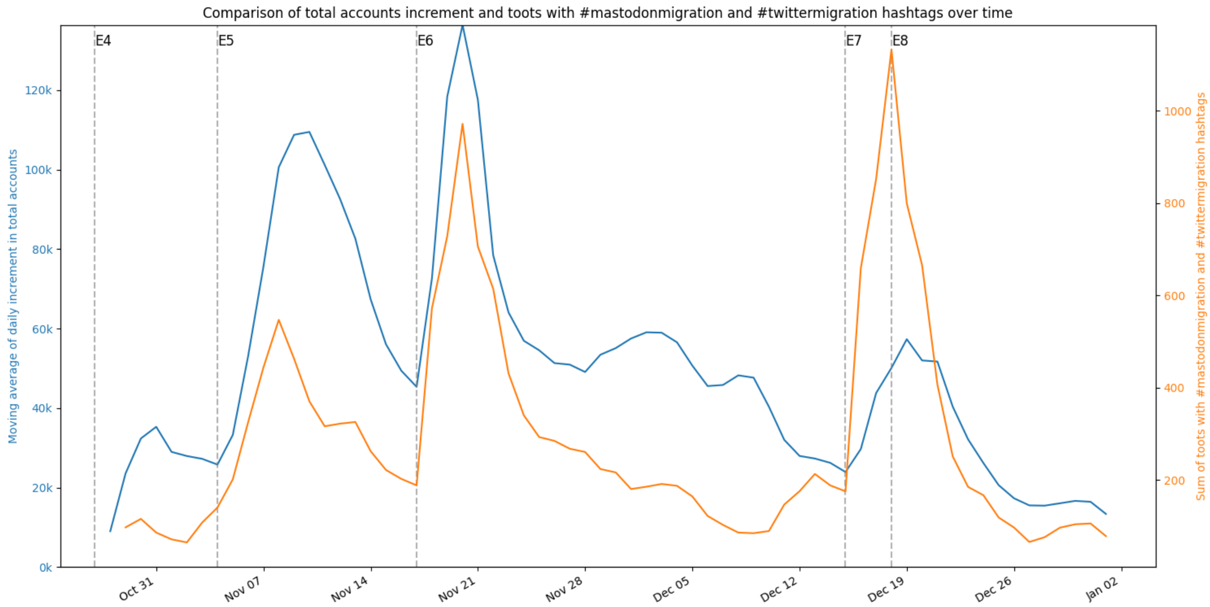

Figure 3: An analysis of the growth in total accounts and toots, in relation to the hashtags *#mastodonmigration* and *#twittermigration*, over time, while considering events associated with Elon Musk (period 2022/10/20 - 2023/01/01).

To improve interpretability, we attempted to correlate topic keywords with the most frequent hashtags, hypothesizing that user-generated labels could provide valuable context [15]. Yet, this did not appreciably enhance topic interpretability. The most frequent hashtags appeared to capture what could be referred to as a *background noise*, reflecting the persistent, macroscopic discourse of the period under investigation, such as ***#twexit***, ***#twitterexodus***, ***#elonmusk***, and so forth. This suggests that these prevalent hashtags do not necessarily coincide with the nuanced themes within the toots, but instead encapsulate the larger, more enduring narratives.

These results suggest that the LDA-detected topics might be too granular and specific, failing to align with broader, macroscopic narratives encapsulated by the frequent hashtags. Alternatively, it could be that these user-generated hashtags do not precisely mirror the nuanced, latent themes within the toots [18]. The observed background noise in the form of persistent, high-frequency hashtags, exemplifies this disconnect. Further research is necessary to fully understand and bridge this discrepancy.

5 CONCLUSION

Our study investigated the impact of Elon Musk's acquisition of Twitter on user migration and engagement on the Mastodon social media platform. We found significant correlations between Musk-related events and spikes in new user registrations on Mastodon. Our analysis also revealed high levels of initial engagement by new users, although it sensibly gradually declined in the days subsequent to enrollment, demonstrating Mastodon's capability to retain users over time.

Despite the limitations in our thematic analysis of discourse on Mastodon, our findings underscore the profound influence of corporate actions and powerful individuals on public digital behavior. As we move forward, it's crucial to conduct further investigations, possibly involving other social media platforms and events. Future research should aim to shed light on how power dynamics in the digital sphere influence user behavior and migration patterns, and the societal implications of these phenomena. This research opens new horizons for understanding the interplay between high-level corporate decisions, the actions of influential individuals, and their cascading effects on digital communities.

REFERENCES

[1] Kenneth C. Anaobi, Onyeka Emebo, and Chijioke Udeze. 2023. Mastodon: The Resilient Social Network. *Social Network Analysis and Mining* 13, 1 (2023).

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[3] Niels Brugger and Ian Milligan. 2019. The Archived Web for Humanities and Social Sciences Research. *Routledge Companion to Media Studies and Digital Humanities* (2019).

[4] Tim Chambers. 2022. Snapshot of the #TwitterMigration. https://www.deweysquare.com/wp-content/uploads/2022/12/DSG-Snapshot-of-the-Twitter-Migration-December-12-2022.pdf. December 2022 Report - Dewey Square Group.

[5] Tim Chambers. 2023. Snapshot of the #TwitterMigration. http://www.deweysquare.com/wp-content/uploads/2023/06/DSG-Snapshot-of-the-Twitter-Migration-June-2023.pdf. June 2023 Report - Dewey Square Group.

[6] Tim Chambers. 2023. Snapshot of the #TwitterMigration. https://www.deweysquare.com/wp-content/uploads/2023/03/DSG-Snapshot-of-the-Twitter-Migration-March-2023.pdf. March 2023 Report - Dewey Square Group.

[7] Anatoliy Gruzd and Philip Mai. 2021. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. *Big Data & Society* 8, 1 (2021).

[8] Steven LaCava and Xueting Xiao. 2021. Migration to Decentralized Social Media: The Case of CampusWire. *First Monday* 26, 3 (2021).

[9] Christian Marazzi. 2002. *Capitale e linguaggio*. DeriveApprodi, Rome.

Table 1: Chronology of events related to Musk's Twitter acquisition

No.	Date	Event
1	2022-04-14	Elon Musk proposes an unsolicited, non-binding purchase of Twitter for $43 billion, or $54.20 per share, with intentions to privatize the company.
2	2022-04-25	Twitter's board unilaterally approves Musk's buyout offer valued at $44 billion.
3	2022-09-13	A court ruling rejects Musk's effort to delay the trial over legal disputes connected to Twitter's acquisition.
4	2022-10-27	The acquisition deal reaches completion, positioning Musk as the new owner and CEO of Twitter; the company privatizes and folds into a new parent company, X Corp.
5	2022-11-04	Musk declares a reduction in Twitter's workforce by half as part of a cost-cutting initiative.
6	2022-11-17	Musk confronts employees with an ultimatum: commit to a demanding, engineering-focused work culture or depart with severance pay, leading some employees to contemplate a mass resignation.
7	2022-12-15	Elon Musk bars several tech journalists, including correspondents from The New York Times, The Washington Post, and CNN, for allegedly violating Twitter's newly enacted location-revealing rules.
8	2022-12-18	Twitter introduces a *promotion of alternative social platforms policy*, which threatens to suspend or remove accounts promoting their handles from other platforms such as Facebook, Instagram, Mastodon, and others.
9	2023-02-14	Elon Musk instructs engineers to launch a new algorithm that amplifies the visibility of Musk's tweets a thousandfold, ensuring over 90% of Musk's followers see them.
10	2023-03-19	As the feud with journalists escalates, Elon Musk introduces a poop emoji auto-response for press emails.

Table 2: Analysis of topics and hashtags

Event date	LDA topic	Most frequent hashtags
2022-10-27	rise, paolo, bi, fields, italian, count, goal, wine, encouraged, newusers	introduction, mastodon, fediverse, introductions, twitter
2022-11-04	harassing, still, mit, joked, estebanmoro, sickening, hairy, warped, traffic, tenforward]	introduction, mastodon, twitter, fediverse, twexit
2022-11-17	dreamtraveler247colors, yesterday, choosing, elonmusktakeover, halo, arduous, axebopzzptdoadbvgqg, toujours, sink, dkgreen	mastodon, twitter, riptwitter, introduction, twitterexodus
2022-12-15	devus, yesterday, porch, deplatform, japaneselanguage, landscape, antifa, spirituality, nfl, borrowed	twitter, mastodon, twitterexodus, elonmusk, newhere
2022-12-18	cw, myself, animalcruelty, mars, propagated, prepare, lightning, style, challenge, imitate]	twitter, mastodon, twitterexodus, elonmusk, birdsite

Note: This table provides a snapshot of the Latent Dirichlet Allocation (LDA) topic modelling and hashtag frequency analysis. It does not encompass the entirety of the data but offers a representative subset.

[10] Mastodon 2023. *Mastodon documentation: Overview.* Mastodon. https://docs.joinmastodon.org/

[11] Dharini Ramachandran and R Parvathi. 2019. Analysis of Twitter Specific Preprocessing Technique for Tweets. *Procedia Computer Science* 165 (2019), 245–251. https://doi.org/10.1016/j.procs.2020.01.083 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION, 2019 November 11-12, 2019.

[12] Bhuvana Raman and Shweta Sharma. 2019. Decentralized Online Social Networks: A Comparative Analysis. *2019 IEEE International Conference on Decentralized Applications and Infrastructures* (2019).

[13] Frances Shaw. 2020. Digital Activism, Hashtags and the Public Sphere. *The Wiley Blackwell Encyclopedia of Sociology* (2020).

[14] Symeon Symeonidis, Dimitrios Effrosynidis, and Avi Arampatzis. 2018. A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Systems with Applications* 110 (2018), 298–310. https://doi.org/10.1016/j.eswa.2018.06.022

[15] Ming-Hsiang Tsou, Chin-Te Jung, Christopher Allen, Jiue-An Yang, Su Yeon Han, Brian Spitzberg, and Jeanne Dozier. 2015. Data-driven techniques in disaster information twitter data mining. In *ISCRAM*.

[16] Artemi Vazdekis, Wainer Lusoli, and Adrián A. Díaz-Faes. 2020. Distributed Digital Activism: The Dynamics of Content Creation and Sharing on Mastodon. *Policy & Internet* 12, 1 (2020).

[17] Katrin Weller and Katharina Kinder-Kurlanda. 2018. A Manifesto for Data Sharing in Social Media Research. *Proceedings of the 9th International Conference on Social Media and Society* (2018).

[18] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing twitter and traditional media using topic models. In *European conference on information retrieval*. Springer, 338–349.

[19] Umar Zia and Faisal A. Butt. 2023. Federated Social Networks: A Case Study of Mastodon. *International Journal of Web Based Communities* 19, 1 (2023).

[20] Matteo Zignani, Sabrina Gaito, and Gian Paolo Rossi. 2018. The Life and Death of Online Social Networks: An Empirical Analysis of Mastodon. *11th International AAAI Conference on Web and Social Media* (2018).

[21] Matteo Zignani, Sabrina Gaito, and Gian Paolo Rossi. 2019. Follow the "Mastodon": Structure and Evolution of a Decentralized Online Social Network. *12th International AAAI Conference on Web and Social Media* (2019).

[22] Danielle Zulli. 2020. Exploring Engagement on Mastodon Through the Lens of Information Network Subcultures. *First Monday* 25, 11 (2020).