MASTERARBEIT | MASTER'S THESIS

Titel | Title

Sustaining Decentralised Social Platforms: Analysing User Activity and Governance Structures on Mastodon

verfasst von | submitted by

Stefanie Kogler BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 926
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Wirtschaftsinformatik
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Math. Dr. Peter Reichl Privatdoz.

Abstrakt

Die rasante Entwicklung sozialer Netzwerke hat die Art und Weise verändert, wie Menschen miteinander interagieren. Bedenken hinsichtlich der Privatsphäre, des Datenschutzes und des Einflusses großer Unternehmen haben jedoch zur Entstehung dezentraler Alternativen geführt. Im Gegensatz zu kommerziellen Plattformen versprechen dezentrale soziale Netzwerke eine größere Autonomie für Nutzende und eine Community-basierte Verwaltung. Diese Masterarbeit untersucht die Aktivität auf Mastodon, einem der bekanntesten dezentralen sozialen Netzwerke, und die Faktoren, die das Verhalten der Nutzenden beeinflussen, darunter Verwaltungsstrukturen, technische Infrastruktur und Engagement. Mithilfe eines quantitativen Ansatzes kombiniert diese Studie die Analyse von Mastodon-API-Daten mit einer groß angelegten Umfrage unter Nutzenden. Die Ergebnisse zeigen, dass die Größe der Instanz und aktives Engagement die stärksten Prädiktoren für Aktivität sind, während die technische Infrastruktur eher eine unterstützende Rolle spielt. Administrative Strukturen wie Moderationspraktiken und Community-Richtlinien weisen eine schwächere, aber positive Korrelation mit der Aktivität auf. Dies verdeutlicht die Bedeutung wertvoller Interaktionen und überschaubarer Instanzen, um die Aktivität der Nutzenden auf Mastodon zu fördern. Da dezentrale Plattformen weiter wachsen, bildet diese Studie eine Grundlage für zukünftige Forschung zur Bindung von Nutzenden, Community-Strukturen und der langfristigen Nachhaltigkeit dezentraler sozialer Netzwerke.

Schlagwörter: Dezentrale Soziale Netzwerke, Soziale Medien, Mastodon, Nutzeraktivität, Nachhaltigkeit sozialer Netzwerke

Abstract

The rapid development of online social networks has transformed how people interact and share their thoughts, but growing concerns over privacy, data ownership, and corporate influence have driven a shift toward decentralised alternatives. Unlike commercial platforms, decentralised online social networks (DOSN) promise greater user autonomy and community-driven governance. This study explores user activity on Mastodon, one of the most prominent DOSNs, focusing on key factors that drive user activity, including governance structures, technical infrastructure, and engagement. Using a quantitative approach, this research combines computational analysis of Mastodon API data with a large-scale user survey to examine patterns of user activity. The findings reveal that instance size and active engagement—such as frequent posting and interacting with others—are the strongest predictors of user activity, while technical infrastructure plays a more supportive role rather than a determining one. Governance structures, such as moderation practices and community guidelines, show a weaker but positive correlation with user activity. These insights highlight the importance of fostering meaningful interactions and keeping communities at a manageable size to sustain user activity on Mastodon. As decentralised platforms continue to grow, this research offers a foundation for future studies on user retention, community governance, and the long-term sustainability of DOSNs.

Keywords: Decentralised Online Social Networks, Social Media, Mastodon, User Activity, Social Network Sustainability

Contents

List of Figures

List of Tables

1 Introduction

By the turn of the century, the internet had developed rapidly and with it came a radical transformation in the economic and social life of society. In response to these changes, online social networks emerged, bringing forth new possibilities for communication [19]. Since then, social networks have been growing at a rapid pace, and by 2024, the number of social media users reached 4.95 billion worldwide [27]. From the initial goal to foster communication between groups and individuals, they emerged as a global means for people to interact, share their personal thoughts or express their opinions about political and economic matters [18]. Moreover, they provide the predominant infrastructure to facilitate inter-personal relationships via the Internet and allow users to communicate over large distances instantaneously. This surge in popularity and expansion predominantly favours centralised platforms, where a single entity accumulates a large amount of attention and users [6].

By the end of the year 2023, Facebook recorded over 3 billion monthly active users, demonstrating an upward trend in its global users for more than a decade [26]. X (formerly known as Twitter), a social media platform designed as micro-blogging site, allows users to participate in real-time discussions and explore tweets—a term that has emerged in pop culture, denoting concise postings limited to a specific character count—on their newsfeeds. By 2024, the platform has reached 415 million monthly active users worldwide [80]. The large user base of online social networks generates vast amounts of communication data, offering large-scale insights into the structure of social networks. This has made the use of social network structures for a variety of purposes attractive [19, 4].

As a consequence, there has been a transition from a user-centred to a more profit-driven approach, based on marketing objectives and advertisement mechanisms [6]. The marketing industry has extensively utilised the social network structure, and by 2024, the social media advertising market was estimated to reach 234.14 billion U.S. dollars [25]. Furthermore, centralised platforms have exhibited a lack of regard for user privacy, authority abuse, conflicts of interest, and vendor lock-in. Hence, online users began seeking alternatives to current online social networks, aiming to regain control over their data and limit the influence of major service providers on their personal information. In response to these privacy concerns, both the academic community and the Free and Open Source Software movement concluded that the development of various alternative systems is necessary [7, 83]. Hence, decentralised online social networks (DOSNs) emerged. Essentially, DOSNs are online social platforms built on a distributed system, providing all the functionalities as centralised networks without the disadvantages of a single service provider [34]. Mastodon is a prominent example of a DOSN and is considered as one of the most popular decentralised federated social networks to date. It operates as an open-source micro-blogging platform, utilising the ActivityPub protocol to connect to various instances [61]. The number of registered users on Mastodon has been rapidly increasing over the last few years, however, the number of active users is decreasing [49]. This phenomenon of initial growth and success has been observed in previous studies. However, many fail to sustain user activity, resulting in the transformation of initially active communities into "cyber ghost towns" [63]. To illustrate the decline in active users on Mastodon, a graph created using Python is presented in Figure 1. This graph utilises data from April to August 2024 sourced from the official Mastodon API [58]. It displays the number of servers, registered users, and active users. Despite the growing number of registered users and servers, there is a noticeable decline in active users. Moreover, studies monitoring migration from Twitter to Mastodon in October 2022 have demonstrated that the initial surge in registrations did not result in sustainable user engagement. Individuals who persisted in posting on Mastodon received less interaction than they did on Twitter, which is crucial for maintaining professional discourse [85].

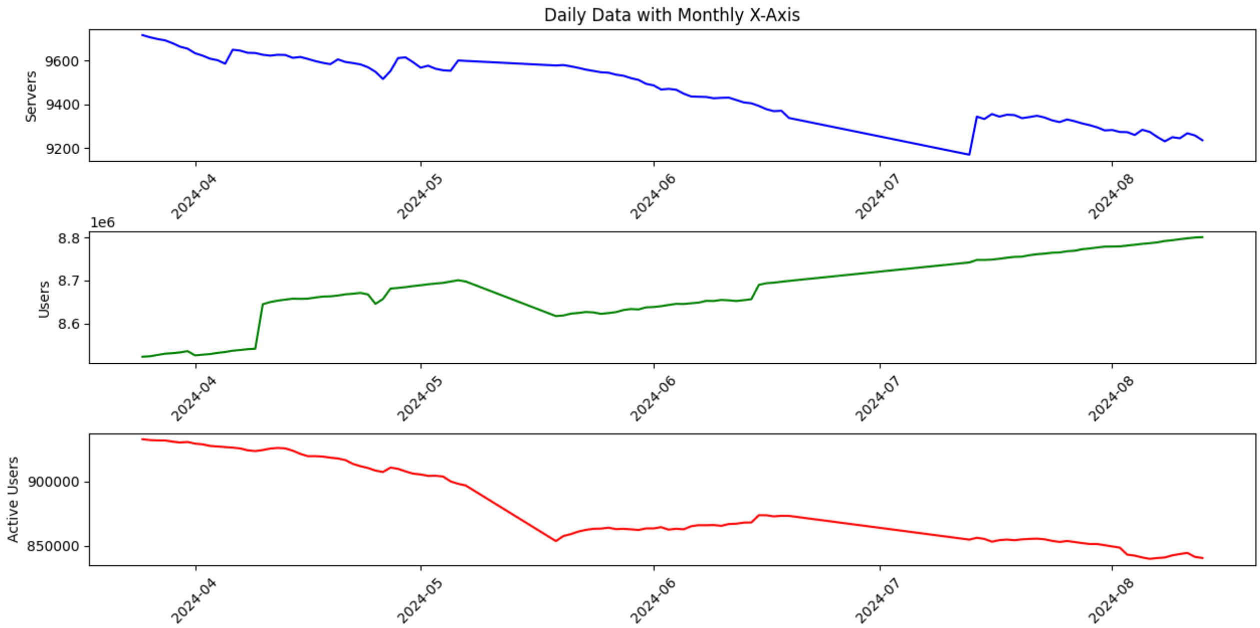

Figure 1: Number of servers, registered and active users on Mastodon

When users register on Mastodon, 96 per cent of them tend to join the top 25 per cent of the largest instances. However, this user-driven pressure towards centralization is counterbalanced by higher user activity on smaller instances [39]. The same study by He et al. (2023) found that users of single-user instances post, on average, 121 per cent more statuses than users on larger instances and that they have in general a more dedicated and proactive user base. On the other side, a small number of instance owners and admins have an impact on a large fraction of new users [39]. Research on another popular Fediverse platform, Pleroma, suggests that some administrators may become overwhelmed by the increasing number of posts and users they need to manage. Signs of this burden include long delays in applying policies against newly federated instances and the fact that only 3.5 per cent of instances distribute the workload across multiple moderators [6]. This imbalance in user activity and engagement has the potential to impact the long-term viability of Mastodon instances. Furthermore, research indicates that if users do not find suitable alternatives, they are likely to revert to established platforms where they have already established communities [85].

2 Online Social Networking Sites

2.1 Definitions

The emergence of social media traces back to the first recorded usage of the term in 1994 on Matisse, a online media environment from Tokyo. In these early days of the commercial Internet, the first social media platforms were developed and launched. [3]. Three years later in 1997, the first recognisable social networking site, SixDegrees.com, was launched [31]. Whereas SixDegrees was not sustainable and had to close down, many other social networking sites [11] followed over time, establishing themselves as one of the most important applications of the Internet [3]. These social media platforms keep attracting millions of users who utilise them every day [14] and are spending several hours on average online [79]. Boyd and Ellison [14] define social networking sites as web-based services, characterised by three distinct features:

1. They allow individuals to construct a public or semi-public profile within a bounded system.
2. Within this system, users may create a list of other users they're connected with.
3. Users can see and navigate through their own list of connections as well as those made by others within the system.

Before 2010, social media was mostly used to connect people with similar interests. For instance, in 2007, Boyd and Ellison [14] introduced the term “social network sites.” They pointed out that the term did not come from the idea of “networking,” since many users were not looking to meet new people. Instead, they were more focused on staying in touch with those they already knew. Thus, when the term "social network sites" was first introduced, it highlighted the feature of managing existing social connections [79]. However, since then the social media landscape has rapidly changed and developed. After 2010, the emphasis shifted to the creation and sharing of user-generated content [3]. This shift can be observed in Bishop’s definition [11], who clearly defines social media as “any online resource with content that is designed to facilitate engagement between individuals”. Moreover, it can be argued that user-generated content is a central component of social media definitions since platforms like YouTube, Facebook, and MySpace primarily focus on cultural and political exchanges. Without user-generated content, these platforms would cease to exist [53].

2.2 History

One of the first recognizable social network sites, SixDegrees.com, allowed users to create profiles, list their friends, and browse through lists of their own friends. However, it was not until 2003 that various new social networking sites began to emerge. These sites predominantly fell into two categories: “profile-centric” platforms focusing on broad audiences or professionals (such as Xing, LinkedIn, or Visible Path), and “passion-centric” platforms aimed at connecting individuals with shared interests, ranging from travellers and activists to church members [14]. With the development of Web 2.0, gatekeepers saw a decentralization of power since virtually everyone was able to produce and share media. Prior to this, mass media predominantly broadcasted from a select few to a broad audience [31]. Websites initially dedicated to media sharing began integrating social networking features, ultimately transforming into social networking sites. Notable examples illustrating this transition include Flickr, a platform for sharing user-generated pictures; Last.FM, which focuses on music; and YouTube, renowned for video-sharing [14]. Nowadays, online social networks are considered as one of the most popular applications on the Internet [9]. As of 2024, daily social media usage has amounted to an average of 143 minutes per day, with Brazil leading at three hours and 49 minutes [79]. Social media platforms have revolutionized how organizations, communities, and individuals communicate and share information with each other [34, 9].

2.3 Challenges of centralised Social Networking Platforms

With the rapid growth of social media, many concerns about centralised social network sites have surfaced. A 2019 study by Statista found that the influence of social media extends beyond online activities, influencing offline behaviour too. While users appreciate the increased access to information and communication, concerns about personal privacy, political polarization, and heightened distractions were raised [79]. When individuals sign up for an online social network, they create profiles containing personal information such as names, social connections, gender and employment statuses. The main objective behind sharing this information is to explore new opportunities or establish new connections [45]. Together with the rapid growth of centralised platforms, this oversharing of information has prompted a shift from a user-centred to a profit-driven approach. As users gradually share more about their lives on social media platforms, this transition has given rise to information bubbles, echo chambers and privacy concerns [45, 62]. Centralised social media platforms alienated users from the surveillance and commercialization of their content in order to maintain the economic system. This was facilitated, for instance, by implementing surveillance software alongside terms of service agreements that grant the platform permission to reuse user information [53]. Consequently, users essentially performed the task of profiling themselves through actions such as liking and tweeting, while site owners monetised the resulting user-generated data by selling it to marketers and advertisers. As a result, the construction of user data profiles and the sale of user attention and data have driven the growth of a highly sophisticated Internet marketing industry [31]. The general public has become increasingly aware of cases, where data from Facebook, have been sold without the consent of its rightful owners. This instance represents only one of numerous legal issues, involving not only Facebook but also other online social networks such as Google+ and Twitter [34]. In addition to the sale of data without consent, centralised social media platforms have been observed to conceal the manner in which they utilise the data of their users. Content and trends emerge without end-users being aware of their origin. This encompasses not only algorithms but also information regarding the storage of personal data, the individuals or entities with access to it, and the inferences drawn from it [94].

Gehl [31] argued that although OSN facilitates broader public involvement in media creation and distribution, dominant platforms like Facebook, Google, and Twitter, still retain and exaggerate certain issues related to the power dynamics of traditional mass media. Moreover, the Web has evolved from many heterogeneous sites to a small number of platforms applying near-monopolistic control over search, content distribution and authentication. Since they are owned by corporate entities, they may show hostility towards alternative ideas, discussions, and organizing, particularly if those ideas challenge their own corporate interests [31]. This has been evidenced in the case of social media platforms shutting down minorities that would require a more nuanced approach to community guidelines and regulations [24, 5]. Moreover, the centralised management of social network sites demonstrated numerous technical and social challenges. From a technical standpoint, issues with performance and growing costs for managing and maintaining infrastructure occurred due to rapid user growth. This manifested in frequent downtime for platforms like Twitter and sluggish performance for others like Facebook [23]. The substantial outage of October 2021 further underscored these challenges, as Facebook, Instagram, and WhatsApp experienced a significant service disruption, thereby revealing the vulnerabilities and widespread impact of such failures on users and businesses alike. This incident highlighted the critical dependence on these platforms and the potential for substantial disruption when they are offline [46]. From the social perspective, the unrestricted sharing of information has resulted in collisions between social circles, as networking sites lacked robust privacy measures. This breach of trust between users and platform providers has played a significant role in the downfall of early social networking services [23].

3 Decentralised Online Social Networking Sites

3.1 Definitions

Currently, centralised social networking sites are experiencing a transformation, urging users to explore new modes of social interaction [52]. Alternative social media platforms have responded to this demand offering users the ability to share content and connect with one another, while at the same time resisting the commercialisation of speech, providing greater access for shaping the underlying technical infrastructure and experimenting in numerous ways to protect personal privacy [31]. A decentralised online social network has been defined as “an online social network implemented on a distributed information management platform” [23]. They are constructed differently from centralised online social networks, as they provide users with access to more than just interfaces [31]. This may entail a network of trusted servers or a peer-to-peer system. Unlike centralised social networks, there is no single provider; instead, a group of peers collectively handle the tasks required to operate the system [23]. It has also been described as a set of nodes that collaborate to ensure all the functionalities provided by a centralised online social network [34]. These systems may be federated or distributed, with “distribution” referring to peer-to-peer technology without the need of central servers [31]. The Federation mitigates the risk of a complete network failure by employing a distributed architecture comprising multiple key nodes, rather than a single central node. In such systems, users establish connections with individual servers that communicate with each other across a larger network [94].

Shifting from a centralised to a distributed architecture provides users with the ability to leverage several distributed models [34]. The shift in the underlying technology leads to different outcomes. From a privacy standpoint, there is no centralised data collection since data is distributed among many autonomous instances, which makes privacy-intrusive data mining more challenging and ultimately leads to a reduction in economic incentives for advertising [23, 66]. Moreover, the absence of a central entity enables higher stability in terms of service agreements. It also allows users to utilise their own storage or cloud storage, facilitate delay-tolerant social networks, and treat local content locally [23]. Finally, DOSNs make data ownership more transparent, and the absence of centralization could improve the system's resilience against legal, technical, or regulatory challenges [66].

A prominent example of an ecosystem of decentralised, interoperable platforms is the Fediverse, which operates on a common underlying protocol, the ActivityPub [89]. In contrast to centralised social media, the Fediverse allows anyone to create a service instance by deploying open-source code on a server [1]. This fosters interaction between instances and consolidates users of decentralised protocols, ultimately resulting in a globally integrated service [66]. Consequently, users can register with these instances and utilise their services. Regardless of the instance or service a Fediverse user is registered with, they can interact with each other. For example, users from Pleroma, a micro-blogging service, can interact with users from another Pleroma instance, as well as with users across other services such as Peertube (video streaming) or Mastodon (a micro-blogging service) [1]. Some of these services provide users with alternatives to major social media platforms like Facebook and Twitter [66].

3.2 ActivityPub

ActivityPub is a decentralised social networking protocol developed by the Social Web Working Group alongside other related protocols like Micropub [88]. The ActivityPub protocol consists of two layers [89]:

- A server-to-server federation protocol that allows distributed websites to share information.
- A client-to-Server Protocol that allows users to engage with ActivityPub servers using different client applications including mobile, desktop, and web apps.

In the ActivityPub protocol, each user, known as "actor", has several key endpoints, including an inbox and an outbox, where messages can be either sent to or received from other actors. These endpoints are outlined in the actor's ActivityStreams description. ActivityStreams, which uses a JSON-based format, includes all the necessary terms to represent the various activities and content shared within a social network. While ActivityStreams already covers a wide range of vocabulary, it can be expanded for specific cases using JSON-LD [87].

There are four key interaction patterns that can be utilised within the ActivityPub protocol:

- POST to Inbox: Sending a message to another actor's inbox, used in server-to-server communication (federation).
- GET from Inbox: Retrieving messages from an actor's inbox, used in client-to-server communication.
- POST to Outbox: Sending messages from an actor's outbox, used in client-to-server communication.
- GET from Outbox: Viewing messages sent by an actor, used in both client-to-server and server-to-server communication.

In a federated system, communication typically occurs when servers post messages from actors to the inboxes of other actors on their respective servers. First, an actor composes a message, which the server recognises as recently created and wraps it in a "Create" task. This is then send to the recipient's inbox. The recipient's server processes the message by looking up the ActivityStreams actor object and finding the inbox endpoint. In the last step, the recipient's server sends the message to the inbox via a POST request. This procedure includes two forms of communication: server-to-server and client-to-server. The message can be retrieved later by the recipient with a GET request to their inbox. Furthermore, actors have the ability to engage with content, such as liking or responding to messages, and then sharing these actions on their outboxes [89].

3.3 Mastodon

This thesis will focus on Mastodon, a decentralised social media platform that has gained considerable popularity among users of alternative social networks.

3.3.1 Overview

Mastodon challenges the prevailing concepts surrounding the definition of social media. In particular, it calls into question the assumptions that: (1) social media is characterised by a centralised logic and administration; (2) social media is most effective when its internal structures are concealed from users; and (3) the primary indicator of social media success is the growth in user numbers [94]. Mastodon is a Ruby on Rails application with its front-end built using React.js [57]. Moreover, it is a free and open-source software project licensed with the GNU Affero General Public License. This licence permits the code to be open for examination and available for anyone to freely alter. Mastodon and its associated documentation and policy statements are developed collaboratively using the GitHub platform. This allows any interested party to view and modify the internal details of networking algorithms, interface design and software-to-hardware relationships [94]. The open-source software permits users to establish their own instances or servers, which can then be accessed and utilised by any individual within a local community [66].

It is operating on ActivityPub [89] and is therefore, part of the broader Fediverse, where individuals can establish and manage their own Mastodon server, also referred to as an 'instance'. These instances can federate with each other, enabling users from one instance to follow another user from another instance. Decentralisation herein allows users to join independent instances without losing the capability of interacting with each other on a global scale [39]. Federation represents the most significant distinction between Mastodon and other centralised social media platforms. This mechanism provides insights into the interplay between technological design and social interactions on alternative social media sites [94].

Despite being released in 2016, Mastodon gained significant attention as a new decentralised social media following Elon Musk's acquisition of Twitter in October 2022 [39]. According to various statistics, both the Fediverse [30] and Mastodon [49] experienced a surge in the number of registered and active users after October 2022 and Mastodon reached over 10 million users as of March 2023 [78].

3.3.2 Key Features and Functionalities of Mastodon

Apart from friends or followers, profiles, comments and private messaging, social media platforms demonstrate significant variation in their user base and features. From photo and video sharing to built-in blogging and instant messaging features, some platforms have integrated a wide range of content-sharing technologies [14]. Mastodon can be defined as micro-blogging platform, which allows publishing brief updates, or "posts," to a stream on a profile. These posts may include text, as well as optional attachments such as images, audio, video, or polls. Additionally, Mastodon allows its users to follow friends and discover new connections [57]. The features of Mastodon are frequently compared to Twitter. In Mastodon, users can create or sign up with an instance and then begin posting "toots." These toots are shared with their own instance followers and, through the federated system, can also be shared with users on remote instances. This setup gave rise to the renowned inter-domain (federated) model, akin to the inter-domain model of the email system [66]. After logging in, the Mastodon interface is shows the user three different timelines:

- a home timeline, showing statuses shared by accounts followed by the user,
- a local timeline, displaying statuses from the same instance and
- a federated timeline, featuring all statuses retrieved from remote instances [39].

These features collectively provide a dynamic and interconnected user experience, thereby enabling seamless communication and content sharing across different user communities within the Mastodon network.

3.3.3 Challenges and Limitations of Mastodon

Despite offering a compelling alternative to centralised social networking platforms, Mastodon is not without limitations. Some of these limitations are listed here as follows:

1. There is a tendency towards centralisation. Mastodon is a decentralised platform comprising thousands of independent instances. However, there is a discernible trend towards centralisation, with a significant number of users registering on a limited set of instances. One illustrative example is mastodon.social, an instance operated by Mastodon GmbH, which accounts for 20 per cent of migrated users. It has been found that 96 per cent of users join the top 25 per cent of largest instances [39].

2. Content Moderation presents a significant challenge for moderators and administrators of decentralised online social networks. In addition to handling content produced internally, these individuals must also filter and manage relationships with other instances. [13]. This can result in an overwhelming workload, particularly for moderators on larger instances who lack sufficient resources. This is evidenced by the extended delays in implementing policies against newly federated instances [6].

3. Decentralised platforms are misused for the coordination and dissemination of harmful material. In the absence of a central authority responsible for defining harmful or toxic content, such material can spread rapidly across the network [38]. Related studies on toxicity on Mastodon and Pleroma have tested the efficacy of defederating instances in combating toxicity [22]. However, the instances' toxicity has not diminished, indicating that toxicity remains a significant challenge for decentralised online social networks.

4. At the time of writing, user activity on Mastodon has been declining [58]. Studies indicate that the decentralised structure of Mastodon and competition from alternative platforms, such as Bluesky and Threads, present significant challenges to sustaining user engagement. Following an initial period of enthusiasm, many users reduced their activity levels, and those who remained experienced lower engagement than on Twitter [85].

4 Challenges in Content Moderation and Governance

The regulation of social media platforms and how content is moderated has been a topic of considerable debate among legal and social academic scholars for some time. It has been customary for states to develop regulatory frameworks in accordance with established legal principles. In response to a series of incidents, European states have implemented policies designed to address the issues of hate speech and disinformation [40]. Nevertheless, it is not only states and governments that regulate the free speech on social media platforms; private online platforms also play an increasingly essential role in the participation of democratic culture. A significant number of online platforms proactively manage the content created by their users, thereby influencing the nature of discourse within their own "systems of governance" [50]. When agreeing to a platform's Terms of Service and Community Guidelines, users consent to operate within an ideological system, which might influence the structures of power related to gender, race, class, sexuality, and other social issues [93]. This has had negative consequences for many minority groups. One such group is sex workers, who have seen social media platforms shut down their forums, de-platform them or ban their content. These restrictions on content have significant implications for the ability of sex workers to perform their jobs safely, find community support, or advocate for themselves. In this context, Davisson [24] characterises community guidelines as "culturally situated statements" that have the potential to disproportionately impact stigmatised groups, rather than as neutral documents. Similarly, transgender individuals encounter considerable obstacles as a result of these policies. Legal cases involving transgender individuals demonstrate how ostensibly neutral regulations can lead to the silencing of their voices and restrict critical activities, such as fundraising for essential services [5].

On the other hand side, more liberal community guidelines could result in an incline of hate speech. After Elon Musk's acquisition of Twitter in October 2022, the restrictions on content moderation were relaxed. However, the newly introduced policies resulted in a significant rise in hate speech [41]. The use of centralised social media platforms has sparked controversy due to the widespread presence of offensive content. When confronted with such content, users are typically limited to flagging the material and awaiting a moderator's review and potential removal of the content in question [94]. However, evidence has shown that the removal of content on social media sites is not consistently applied across all user groups. Consequently, there is a risk of disproportionate censorship related to individuals' genders, races, or political orientations [36].

In contrast to centralised platforms, alternative social media such as Mastodon are independently managed. This implies that each Mastodon instance is capable of establishing its own moderation and content regulations [94]. When standards among interconnected, autonomous communities differ without a single controlling entity, the management and moderation of these communities becomes a challenging task [22]. The content moderation capabilities of federated platforms are less developed, robust and scalable than those of their centralised counterparts. The architectural constraints in place limit their capacity to defend against common social media behavioural threats [70]. This lack of content moderation not only hinders the widespread adoption of such platform [33] but has also shown to pose persistent challenges, such as spam and coordinated behaviour [70]. On Mastodon, founding administrators typically monitor the content through a community-based moderation that is closer to the end-users. As a result, there is considerable variation in the level of moderation applied to different instances, with some being lightly moderated and subject to less stringent rules, while others have more tightly regulated content [94]. In the event that a user objects to the choices made by one instance, platform interoperability and account portability permit users to engage with a wide variety of alternatives without having to sacrifice their network or content [70]. This configuration permits users to elect their preferred instance and moderation policy. Should they disagree with the latter, they are at liberty to contact their instance administrator [94]. By shifting governance decisions from a single entity to more localized choices made by moderators and administrators, Mastodon empowers users with greater agency, autonomy, and ownership over their social media experience. [70].

On Mastodon, moderators and administrators take on an important leadership role. However, the distribution of the moderation effort across multiple entities introduces an additional layer of complexity. Each instance comprises a set of moderators and administrators, whose role is to manage and filter content on their respective instance. To address this challenge, Mastodon offers administrators a comprehensive suite of tools, including those that enable the implementation of decentralised moderation policies across multiple instances. Among these are blocklisting mechanisms, which facilitate the definition of rules aimed at preventing interactions with specific instances and, consequently, the propagation of hateful, harmful, and NSFW (not safe for work) content [13]. As this process is not automated, many administrators, especially of larger instances, experience difficulties in allocating sufficient resources to handle the associated moderation overhead [6].

5 Engagement Dynamics on Social Media

This chapter examines user engagement and how it can differ on centralised versus decentralised social media platforms. It is key to make a distinction between active engagement, such as liking, commenting, and sharing, and passive engagement, which is similar to watching TV or listening to the radio [64].

5.1 User Engagement on centralised Social Media Platforms

One aspect that attracts companies to centralised social media platforms is their ability to facilitate direct communication between brands and consumers. These interactions may be likes, comments or shares of the content with their friends and followers. Therefore, social media has been employed as a marketing instrument by brands and marketers alike [8]. In 2013, Brodie et al. [16] initiated the conceptualisation of the term 'engagement' in the context of brand communities. Consequently, the subject of methods to enhance engagement has become a prominent area of research in marketing [8, 56, 76, 73]. One of the key aspects of these studies was the impact of social media content posted by firms. It was demonstrated that there is a distinction in users' engagement behaviour based on the type of platform. Furthermore, the content format exerts a significant influence on the efficacy of social media engagement behaviour. For instance, the content type could be rational, emotional, or transactional, and based on the type of content, the format should be adjusted. One example provided by the study was that emotional content in video format drives active engagement through comments, making it more conducive to fostering interaction [76]. Furthermore, the nature of active engagement underwent a transformation during certain periods of time. For example, during the global SARS-COV-2 pandemic, there was a notable increase in social media traffic. However, active engagement that is linked to positive psychological and social outcomes exhibited a decline [64]. Despite all the research examining the optimal strategies for enhancing the efficacy of content, centralised social media platforms exercise complete control over the social engagement of their respective communities, establishing the rules and regulations that govern them [94].

5.2 User Engagement on Mastodon

Decentralised networks foster sociality through user-driven negotiation and interest-based engagement. This is in stark contrast to the pre-existing technical structures and platform capitalism that shape social interactions on centralised social media networks. The objective of Mastodon is to facilitate interest-based community engagement, and as there are no restrictions on who can create an instance, numerous niche groups have emerged [94]. Despite the wide range of interests, the maintenance of professional discourse is dependent upon a minimum level of engagement. This was evidenced in the study conducted by Wang et al. in 2024 [85], which monitored the large migration of Twitter users to Mastodon. Following an initial surge in sign-ups, user engagement declined and failed to translate into sustainable activity. Users who continued to post did not receive the same level of engagement as on Twitter, which is important for sustaining professional discourse. In the context of Mastodon, engagement would encompass the posting of statuses so called "Toots", the expression of likes, comments or shares of other users' posts. Another study looking at user engagement found that individuals on single user instances tend to post 121 per cent more statuses than those on larger instances. They also demonstrated a more engaged and proactive user base. In fact, some users create their own instance to have more control over the environment. These users were inherently more active [39]. In contrast to centralised social media platforms, where users have limited authority over content, decentralised social media users are empowered to monitor their own and others' engagement according to community values [94]. This can be partially achieved through a mechanism called defederation, which enables the blocking of other servers. The objective is to create a safer environment for decentralised users by blocking toxic content. However, there was no effect on message toxicity and the blocked server showed a decrease in user activity. On the other side the blocking servers did not show any reduction in user activity [22]. Another important point is that facilitating community development is essential to sustain user engagement on decentralised platforms. This can be achieved by reducing the barriers to participation and considering network effects. Achieving these goals requires not only technical solutions, but also the cultivation of communities that can compete with established social media networks and offer comparable user experiences [85].

6 The Significance of User Activity on Social Networking Platforms

The primary indicator of success for large-scale social media platforms is often user growth [94]. For example, in Facebook's quarterly reports, in addition to advertisement revenue, income and other financial highlights, the DAP (daily active people) is reported. In this case DAP refers to the number of unique users who engage with the platform on a daily basis, and it is an important metric for user growth [59]. This is driven mainly by economic value, which results in a narrow focus on growth metrics that do not fully capture the broader social, cultural, and political implications of these platforms [94]. In contrast to this, the study conducted by Zulli et al. in 2020 [94] revealed that, rather than focusing on user growth, Mastodon users and administrators emphasise the number of instances within the Fediverse. Moreover, the researchers identified the number of users, the number of toots and the number of connections with other instances as key metrics for assessing the scale of an instance. This information provides insight into the level of engagement and activity of an instance, enabling potential users to identify instances that align with their social needs. However, it was observed that a specific number of active users is essential for instances to flourish, indicating those who frequently interact with the site. The moderators of these instances demonstrated an acute awareness of the pivotal role played by scale in this context [94].

6.1 Activity Rate: Definition and Measurement Challenges

In the past, there have been different interpretations of user activity. This paper sets out to make a clear distinction between the concept of active users and that of user engagement, which has already been addressed in the preceding chapter. Mastodon defines an active user as a user who has accessed the site or its API within the last 30 days [69]. In contrast, user engagement encompasses users who frequently actively interact with the platform by posting statuses, liking or commenting on other posts.

This thesis examines user activity rate as a dependent variable, which requires two key components: the number of registered users and active users. However, obtaining these figures is challenging in centralised online social networks. In order to estimate the total user base of an online social network, it is necessary to crawl a sample of user profiles. However, there is a lack of ground truth data, which is proprietary data owned by the online service operator [47]. Researchers often have limited or no access to scientifically relevant digital datasets because the companies that operate these digital platforms (e.g. Twitter Inc, Meta Platforms Inc, Google LLC) restrict access technically, legally and financially [28]. As a result, it is challenging to compare activity rates between Mastodon and other platforms.

Facebook, in particular, presents considerable obstacles for large-scale crawling techniques. A study conducted in June 2018 estimated the user base of Facebook to be 2.2 billion, with 1.3 billion individuals accessing the platform daily [47]. This would represent an activity rate of 59 per cent, which is considerably higher than the activity rates observed on Mastodon. However, Facebook is removing "fake accounts" on a daily basis. "Fake account" are accounts with the intention of misleading or deceiving others. These accounts may be created with malicious intent or represent a business, organisation, or non-human entity. In the second quarter of 2024, the fake accounts aggregated to 1.2 billion with a record high of 2.2 billion accounts removed in the first quarter of 2019 [29].

Mastodon, as an open-source platform, lacks a central authority that controls, sells, or restricts access to user data. However, this decentralised structure also means that individual instances apply different criteria for removing inactive profiles, leading to variations in user activity rates. To address this, this thesis employs a multi-method approach to measuring user activity. In the first analysis, the actual user activity rate is calculated by dividing the number of active users by the total number of registered users across Mastodon instances. In the second analysis, a user survey is conducted, in which participants self-report their activity levels on a predefined scale. By combining these two measurement approaches, this study provides a comprehensive overview of user activity patterns on Mastodon.

6.2 Geographic Analysis of Mastodon Activity

In order to gain deeper insights into the demographics of Mastodon users, a list of Mastodon instances from Fediverse Observer [1] has been extracted and visualised with Python. Figure 2 illustrates monthly active users, sorted by country. First, the monthly active users have been subjected to a preprocessing procedure

[1] https://api.Fediverse.observer

with the subtraction of 0.001 per cent from the lower and upper border of the data frame. Secondly, it was necessary to ensure that both the monthly active users and the total registered users had a minimum of one registered and one monthly active user. This process helped to eliminate any potential outliers and ensure that only active instances were included in the data set. It is important to note that deriving the country name from the IP data of an instance was not possible in all cases. A large number of instances have chosen to maintain the confidentiality of this information, and are consequently not represented in Figures 2 and 3, as well as in Table 1.

The preprocessed dataset was used to calculate the total number of registered users in each country and to group them according to country of residence. This approach enabled the identification of the 15 countries with the largest user bases. The next step was to retrieve the monthly active users for these top countries, aggregate the total per country and sort them in descending order. Finally, the activity rate per country was calculated by dividing each country's monthly active users by its total registered users and multiplying by 100 to express it as a percentage. In figure 2 you can see a bar chart demonstrating the countries with the most substantial user bases, ordered according to the number of active users. In October 2024, Germany had the largest number of monthly active users, with a total of over 100,000 active users, followed by the United States, with 77,599 active users, and France, with 75,112. Figure 3 illustrates the activity rates per country. Comparing Figure 3 to Figure 2, it can be observed that countries with larger user bases tend to have lower activity rates. The exception to this seems to be Russia with a discrepancy between the number of non-active users and the number of registered users. Moreover, the 15 countries with the largest user bases indicate a relatively low activity rate, with a mean of 10.58%. When compared to the data of the official Mastodon API, it shows that both outcomes indicate a significant gap between the number of registered users and the number of monthly active users. This issue has been discussed in the introductory chapter [58].

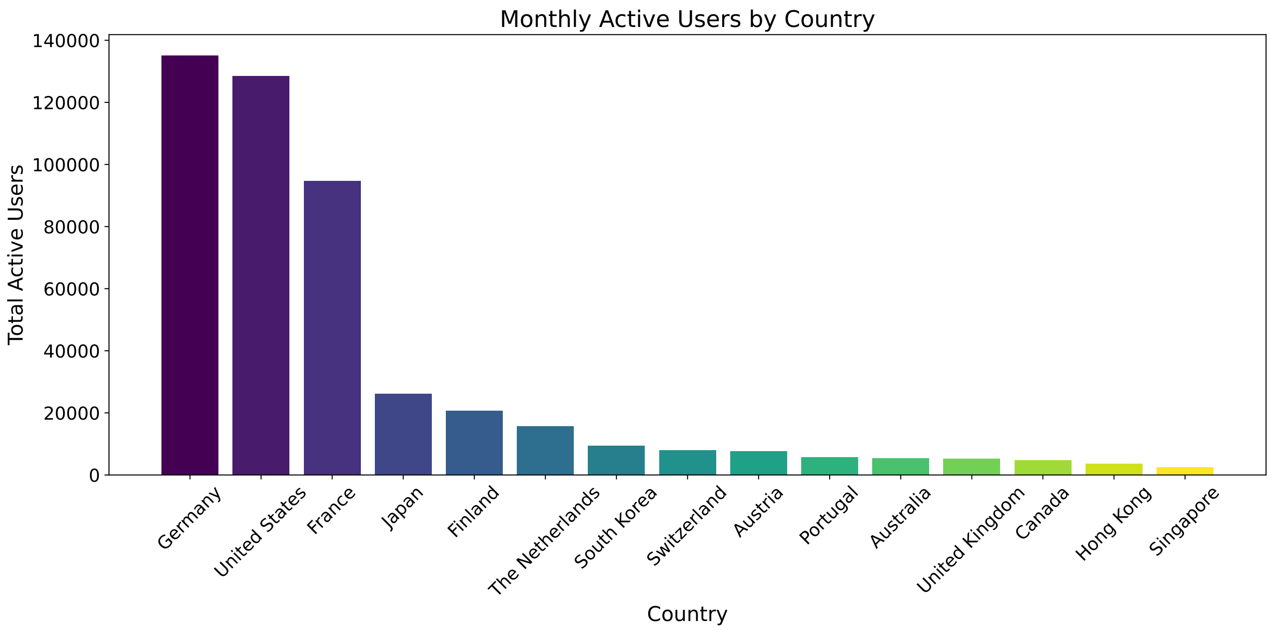

Figure 2: Monthly Active Users by Country

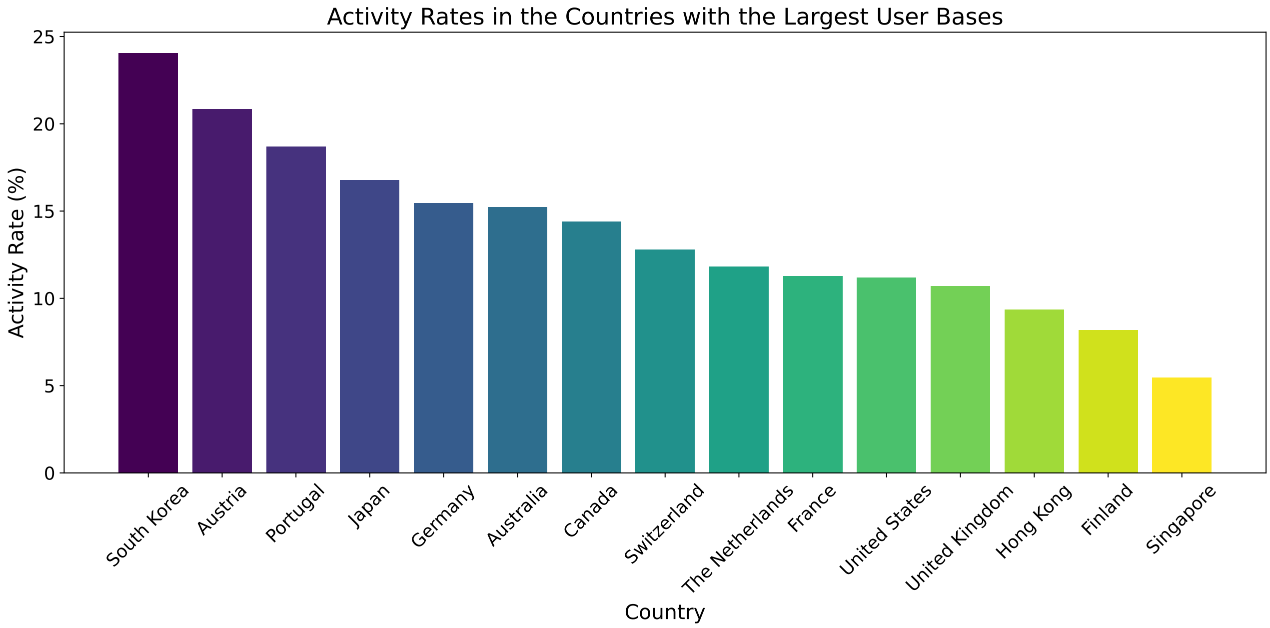

Figure 3: Activity Rates by Country

Table three provides a more detailed description of the activity rate across the countries with the highest number of registered users. The data indicates that the activity rates in Germany (6.22 %), the United States (3.80 %), and Russia (0.21 %) are comparatively low. The highest activity rates can be observed in South Korea (27.76 %), Austria (18.07 %) and Japan (14.83 %). Table 1 shows the countries with the highest amount of registered users. The United States leads the list with a user base exceeding 2 million, while Germany follows with over 1.6 million users. Among 2,045,200 users in the United States, merely 77,676 are active monthly. This again emphasizes the gap between registered users and their actual activity.

Country	**Total Users**	**Total Active Users**	**Activity Rate (%)**
Australia	32,332	3,699	11.44
Austria	29,744	5,376	18.07
Canada	32,474	3,925	12.09
Finland	234,699	20,482	8.73
France	858,385	76,895	8.96
Germany	1,610,006	100,181	6.22
Hong Kong	38,839	3,832	9.87
Japan	128,362	19,041	14.83
Russia	231,781	483	0.21
Singapore	48,963	3,393	6.93
South Korea	37,266	10,345	27.76
Switzerland	80,538	6,508	8.08
The Netherlands	122,262	11,185	9.15
United Kingdom	40,931	5,125	12.52
United States	2,045,200	77,676	3.80

Table 1: The figure highlights the countries with the largest user bases on Mastodon while also illustrating their respective levels of active users and overall activity rates.

7 Key Determinants of User Activity

This chapter explores the key factors that influence user activity by examining various technical, social and infrastructural elements that shape the user experience. Every chapter explores one factor in detail, showing the impact on user behaviour that has been discussed in the literature. Figure 2 provides an overview of the factors that influence user activity. It is hypothesised that technical infrastructure, governance structures and engagement positively correlate with user activity. Furthermore, technical infrastructure comprises several factors, including information overload, security, uptime, latency, sustainability and transparency. Governance structures of a Mastodon instance encompass community guidelines and content moderation.

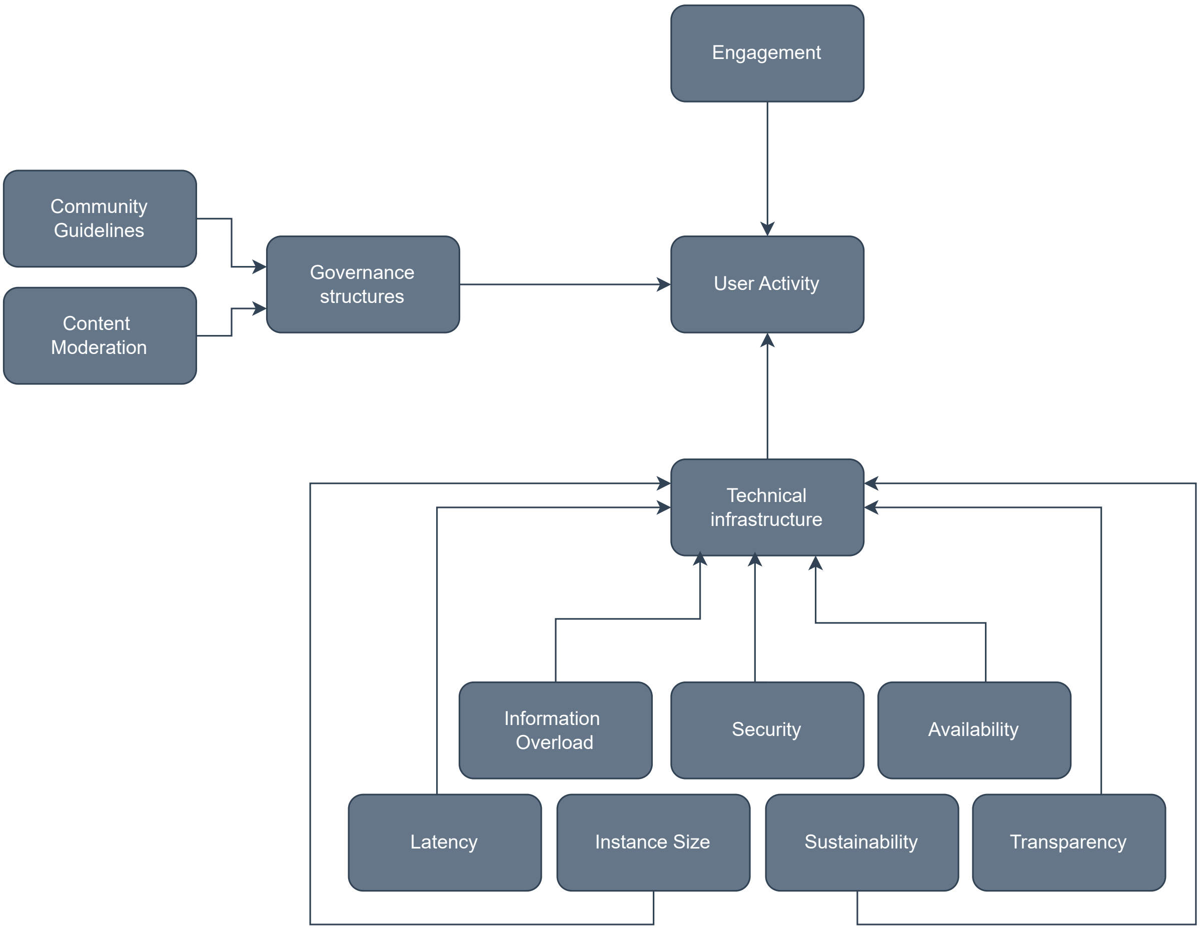

Figure 4: Key factors hypothesised to influence user activity on Mastodon

7.1 Technical infrastructure

The study of Xu et al. in 2014 [92] indicates that user dissatisfaction with technical and information quality is not a significant factor in determining their intention to switch to another system. Nevertheless, in that study users rated the technical and information quality of their social networking sites highly. An earlier study on online community usability found that users prioritize ease of use and system reliability, particularly when seeking and contributing knowledge. These factors were of even greater consequence for those whose objective was to seek knowledge rather than to contribute it [63]. In contrast, Mastodon exhibits a distinctive technological design in comparison to centralised social media platforms [94]. The infrastructure of centralised social media platforms differs from that of decentralised social media, as each instance can be set up using open-source code. Mastodon, for instance, relies heavily on a few large hosting providers, and the regular outages of both autonomous systems and individual instances can cause significant disruption and fragmentation to the network [66]. Consequently, it can be stated that the fundamental basis of centralised social media and decentralised social media is distinct. In light of this, the findings of Xu's study [92] that technical quality is an insignificant factor in user retention should be reevaluated on decentralised platforms. The following chapters will address certain technical factors that may impact user activity and retention.

7.1.1 Information Overload

Prior research has indicated that a number of factors can contribute to feelings of regret regarding the use of social media platforms. This, in turn, may give rise to the intention to cease using them. Among the factors identified are cyberbullying, misinformation, information overload, misuse and online social stress [44]. The phenomenon of information overload has previously been identified as a factor associated with the intention to stop using social media platforms. However, it has also been proposed that advertisement may act as a trigger for that phenomenon [90]. Moreover, algorithms that facilitate the recommendation of content are part of the primary infrastructure on centralised social media platforms. On the social media platform X, content that has been liked by the user is automatically displayed at the top of the feed, thus ensuring that it is seen by a greater number of people [74]. This is achieved through the utilisation of intellectual property regimes and topologies, which serve to conceal the precise functioning of the internal algorithms from the user. In contrast to the aforementioned example, content or trends frequently emerge without any discernible cause or trigger [94]. It is important to note that advertisements and sorting algorithms are absent from decentralised platforms such as Mastodon. This could result in a different outcome when it comes to the impact of information overload on user activity. In this study, information overload is defined as an excessive amount of information on Mastodon that is available to users, which could lead to difficulty in processing the volume or the complexity of the content.

7.1.2 Uptime

In contrast to centralised platforms, which store user data in a single location, decentralised social networking sites largely implement the P2P (peer-to-peer) paradigm, whereby users' devices distribute tasks across multiple devices. While this approach enhances privacy, it also introduces challenges related to data availability, load balancing, and information diffusion. In particular, the issue of data availability represents a significant challenge, especially in the context of user churn, whereby users autonomously disconnect from the decentralised network, which may result in potential data loss. In the event of a user disconnecting, the data stored on the network becomes inaccessible and unavailable. Consequently, the behaviour of participating users has a considerable impact on data availability, with the potential for data to become unavailable or lost [72]. Proposed solutions adopt a user-centric approach to addressing this issue [35], focusing on maintaining data availability even when users are offline. Similarly, Shahriar et al. (2017) [77] demonstrate that leveraging user uptime patterns through structured replication strategies improves data availability and overall network performance. In the context of Mastodon, server uptime reflects the operational accessibility of an instance. Accordingly, this thesis hypothesizes that high uptime positively correlates with user activity.

7.1.3 Security

Social networks carry an inherent risk of causing security problems, including the leakage of confidential information [91]. The findings of a study on decentralised social media indicated that security and privacy breaches constituted a significant obstacle to the widespread adoption and use of decentralised social media platforms [33]. However, the potential for large-scale data breaches is reduced in decentralised social networks relative to their centralised counterparts, given the distributed nature of servers. Administrators of decentralised social networks have the option of hiring data storage services that align with their desired security and protection requirements, thereby enhancing their control over the security of their instance

[48]. Digital certificates are one potential method for enhancing security. Their implementation ensures the security of a connection between a browser and a web server. HTTPS (Hypertext Transfer Protocol Secure), an extension of HTTP, ensures secure connection through the integration of Secure Sockets Layer (SSL) or Transport Layer Security (TLS). Recipients can then ascertain whether a connection is secure by examining the accompanying SSL certificate [54]. On Mastodon, the level of security afforded to each instance can be developed with different degrees of security based on the available time and financial resources.

7.1.4 Latency

The widespread use of social media and other data-heavy applications is putting increasing pressure on existing network infrastructure. Data centres need to handle growing network traffic while also ensuring that performance standards for latency, reliability, flexibility, and scalability are met [37]. A study revealed that for younger generations (ages 16–25), response time is among the most important factors influencing their choice of social networking site (SNS) applications [51]. Latency on Mastodon is associated with the time associated with establishing a connection to the server. This may potentially influence the overall experience of users on the social network.

7.1.5 Sustainability

Environmental sustainability may be a significant factor influencing users' decisions regarding the selection and retention of an instance on Mastodon. Mastodon offers the option of indicating whether the instance is hosted at a green data centre and whether it is part of the Green Web Foundation. The Green Web Foundation is a non-profit organisation whose objective is to achieve a fossil-free internet and to maintain the largest open dataset of websites that run on green energy [32]. On the other side also economic sustainability might be an important factor influencing user activity. A main difference of decentralised and centralised social media is that centralised social media platforms are free to use in exchange for data and content [94]. Mastodon administrators often bear the financial responsibility for maintaining their instance. One potential source of support is Patreon, which enables users to contribute financially to an instance, allowing administrators to invest in additional features and enhancements on Mastodon. [68].

However, these contributions are not typical, and Mastodon's economic structure raises questions about its long-term viability. Without the support of donations, moderators may be less inclined to facilitate instances, which could negatively impact the platform's growth and functionality [94]. This thesis defines sustainability as environmental sustainability, with a particular focus on how users perceive the environmental impact of a Mastodon instance. In contrast economic sustainability focuses on the perspective of the instance's administrators. Therefore, environmental sustainability on Mastodon can be understood as a measure of the extent to which an instance is 'green' or environmentally sustainable, as perceived by Mastodon users.

7.1.6 Transparency

"Transparency is a prerequisite for accountability and responsibility, it can enlighten and therefore empower." [60] Privacy concerns represent one of the most prevalent issues associated with centralised online social networks that has been frequently mentioned in literature [34, 45, 62, 20]. Users often unknowingly disclose their private information [62], which is subsequently stored in centralised servers owned by the companies operating these social media platforms. This social data has been then utilised for commercial and marketing purposes [34, 31], attracting venture capitalists who have identified a new market to exploit and invest in pursuit of revenues and public offerings of stock. Consequently, prominent platforms such as Facebook, Google, and Twitter have encountered pressures from the political economy of the Internet. In contrast, decentralised social media platforms, afford users the option of sharing content while simultaneously exerting greater control over the underlying technical infrastructure and ensuring personal privacy [31]. This thesis defines transparency as how Mastodon administrators and owners communicate their data practices and policies to users, thereby ensuring clear and accessible information regarding the handling of user data. In particular, Mastodon provides the option for instance administrators to link to privacy policies, thereby enabling each instance to offer transparency about its data practices.

7.1.7 Instance Size

Upon joining Mastodon, 96 per cent of users register with the top 25 per cent of the largest instances. However, these larger instances also tend to exhibit lower user activity than smaller instances. [39]. Furthermore, moderators on larger instances face challenges in allocating sufficient resources to an overwhelming workload, which is evidenced by extended delays in implementing policies [6]. As also our study on countries in figure 3 have already shown is that countries with larger user bases tend to have lower activity rates. A considerable number of moderators in smaller instances seek to restrict their communities in order to guarantee the quality of interactions and the efficacy of moderation. To illustrate this, the moderator of Mastodon.at considers a community comprising between 500 and 1,000 users to be optimal for engagement. Notwithstanding the preference for small-scale instances, users benefit from the global reach of the Fediverse, which enables access to diverse content across instances [94]. In light of the aforementioned considerations, it can be posited that instances with a smaller user base tend to exhibit higher levels of activity.

7.2 Governance Structures

Governance structures represent a central aspect of social networking sites. In contrast to other platforms, Mastodon lacks a centralised community guidelines and policies system, with a limited number of moderators tasked with enforcing these guidelines [6]. The right management of the instances and their content could be a crucial factor in the long-term viability Mastodon instances. Therefore, it would be crucial to examine how the role of community guidelines and content moderation is shaping the user experience on Mastodon.

7.2.1 Community Guidelines

This thesis suggests that implementing transparent, inclusive and clear community guidelines and policies may effectively address core issues related to user activity and community building on decentralised platforms. Despite the findings of Xu et al. [92] that moderation policies have limited effect on users' intentions to switch platforms, recent events have demonstrated otherwise. Following the change in Twitter's ownership and alterations to its content moderation policies, a significant number of users sought alternative platforms for their online discourse [85]. Therefore, it would be crucial to investigate the significance of community guidelines on Mastodon and to what extent they could potentially lead to unsustainable instances or user inactivity.

7.2.2 Content Moderation

In a study conducted in 2014 on user retention and acquisition in social networking services, it was found that users may be inclined to consider switching to another platform if they are dissatisfied with the level of socialisation support available to them or if influenced by their peers. This emphasises that the formation of robust community ties represents an effective strategy for the retention of users [92]. Supporting the importance of content moderation, an experiment conducted by Aguerri [2] examined how different presentations of banned discursive behaviour influenced perceptions of content removal. The findings revealed a high level of agreement on the need to remove such content, regardless of how it was presented, and highlighted that the behaviour in question was perceived as serious due to the presumed harm it causes. Building on this, the presence of engaging, experienced moderators who enforce community policies and rules has been demonstrated to have a positive effect on the perceived sociability of an online community [63].

It was also found that the effectiveness of leadership styles in maintaining user engagement in online communities is of significant importance. A diverse range of leaders is needed to address the heterogeneous needs and interests of all participants, which is crucial for sustaining the community's engagement over time [55]. Therefore, it is essential to ensure that Mastodon instances are managed in a way that fosters community-building. Wang et al. [85] proposes that this may be achieved by minimising friction and taking into account network effects with a view to retaining user engagement. This thesis defines content moderation as the active interaction of moderators on Mastodon who enforce community guidelines. This encompasses the removal of harmful content, the issuance of warnings to users, or the blocking of instances that violate community standards.

7.3 Engagement

User engagement, such as likes, comments and shares, serves as a critical form of feedback on social media platforms. This feedback can trigger complex social dynamics, influencing user behaviour in various ways. Research has shown that negative feedback often leads users to alter their posting behaviour, while positive feedback encourages them to maintain their current activity patterns. Conversely, no feedback can result in users becoming disengaged or even leaving the community entirely [21]. Another study in 2024, a decade later, found that little to no interaction on your posts not only indicates stress and negative emotions, but also low self-esteem [84].

Centralised social media design and algorithm choices that are frequently made favour passive social media consumption, as they tend to curate endless feeds of content to keep users scrolling. In order to establish active engagement and interaction, it would be beneficial to incorporate features that promote engagement and provide users with greater control over their social media experience [64]. Moreover, a certain level of engagement is essential for fostering sustained professional discourse on social networking sites [85]. In this thesis, engagement is defined as the active interaction with other users' content. This may include actions such as liking (favouring), commenting on (replying), or sharing the content of other users (boosting). It is proposed that increased engagement not only enhances the quality of social interactions but also directly contributes to higher user activity.

8 Methodology

8.1 Research Questions and Hypotheses

This thesis employed a quantitative approach to examine the relationship between various factors and user activity on Mastodon. Before collecting primary resources, a thorough literature review of statistical information, scientific and academic resources has been conducted. Based on the literature review and the research aims, the first research question can be articulated:

- RQ1: How do governance structures impact user activity on Mastodon?

The literature review gave an overview about the governance structures on Mastodon and outlined the challenges that are inherent to both centralised and decentralised social media. The governance structures that this study focuses on are community guidelines and content moderation. Community guidelines refer to the policies that moderators and administrators can establish on their instance, whereas moderation practices ensure that these guidelines are followed. Community guidelines can foster a safe environment that encourages participation and respectful and safe interactions. Content moderation, on the other hand, addresses inappropriate content, responds to user reports or monitors discussions on their instance. Based on the first research question the following hypothesis can be formulated:

- H1a: Instances with well-defined, transparent and inclusive community guidelines show higher user activity rates.

- H1b: Instances with responsive and proactive content moderation demonstrate higher user activity rates.

The first research question and the respective hypotheses aim to identify the perceived effectiveness of governance structures in promoting user activity. This includes insights into user perceptions of community guidelines and content moderation practices. To answer the first research question, a quantitative survey of Mastodon users was conducted. This survey inquired about the governance structures of their instances and whether they perceive them to be transparent and clear (community guidelines) and proactive (moderation practices). Further details of the quantitative study can be found in the chapter of survey design. Therefore, it is hypothesised that instances with transparent, inclusive and clearly-defined community guidelines and proactive moderation practices should demonstrate higher activity levels. The second research question can be posed as follows:

- RQ2: How does the technical infrastructure affect user activity on Mastodon instances?

The second research question examined the relationship of technical infrastructure with user activity levels on Mastodon instances. The technical infrastructure includes information overload, security, uptime, latency, sustainability, transparency and instance size. These factors were investigated through a computational study with API (application programming interface) data. Therefore, a list of Mastodon instances from an open-source API was retrieved, and the aggregated and processed data were analysed. This resulted in the identification of several interdependences between the predefined factors and the user activity levels. The hypotheses of the second research question are therefore formulated as follows:

- H2a: Instances with higher uptime exhibit higher user activity levels.
- H2b: Instances with lower latency demonstrate higher user activity levels.
- H2c: Instances with enhanced security measures exhibit higher user activity levels.
- H2d: Instances with greater transparency, including clear privacy policies and terms of service, foster higher user activity levels.
- H2e: Information overload leads to lower user activity levels.
- H2f: Environmentally sustainable instances have higher user activity levels.
- H2g: Smaller instances exhibit higher user activity levels.

A robust technical infrastructure is essential for accessing services on a platform. It was hypothesised that instances with greater up-time, more security implementations, transparency, sustainability and lower latency would have higher levels of user activity. A further element of the technical infrastructure is information overload. This factor, related to the user's perceived experience, was evaluated differently through a quantitative user survey. It was hypothesised that information overload should have a negative impact on user activity. The final aspect of technical infrastructure was instance size. It was hypothesised that instances with a smaller user base were to exhibit higher user activity levels. This is an important component of the technical infrastructure, as administrators of an instance have the option to close up registrations and maintain a smaller instance size, which may result in higher user activity. Lastly, the third research question is presented:

- RQ3: How does active user engagement influence user activity on Mastodon?

The objective of the third research question was to determine whether higher levels of engagement correlate with higher user activity levels. To answer the third question, data was collected via a user survey on Mastodon. It was hypothesized that there is a positive correlation between the engagement a user receives and their activity levels. Additionally, it was proposed that frequently engaging with content from other Mastodon users also correlates with higher user activity. Finally, it was suggested that users who toot more frequently were more likely to demonstrate higher user activity.

- H3a: Users who proactively toot demonstrate higher levels of user activity.
- H3b: Users who engage with the content of other users exhibit higher levels of user activity.
- H3c: Users who receive higher engagement on their toots exhibit higher levels of user activity.

Given the limited scope of this study, the focus was on government structures, engagement and technical infrastructure. These categories were chosen based on literature and their importance in a decentralised network. By focusing on the deriving factors, the study aimed to provide deeper insights into user behaviour and user activity dynamics on Mastodon.

8.2 Research design

8.2.1 Quantitative Survey

This thesis is an explanatory, empirical study. The objective of an explanatory study is to test the validity of previously established hypotheses, which are derived from existing theories. In particular, cause-effect relationships are to be confirmed or rejected [28]. In this thesis, the cause would be various predefined factors, such as the technical infrastructure, governance structure or engagement that are hypothesised to influence user behaviour. The effect would be the resulting changes in user activity on Mastodon, described by monthly active users. Explanatory studies are typically fully structured quantitative studies that test precisely the postulated effects. They focus on the detection of effects, but also contribute to the understanding of causal mechanisms in the form of experimental and quasi-experimental studies [28]. In research we differentiate between qualitative and quantitive approaches. Qualitative research aims to gain insight into human behaviour by understanding it. Understanding involves empathizing with human subjects and gaining insight into their personal experiences. This approach often employs case studies that do not allow for generalisations to be made. The emphasis placed on interpretation over objectivity results in perceptions of subjectivity. In contrast, quantitative research is designed to explain human behaviour by developing generalisable rules. It involves controlled conditions and empirical testing of hypotheses, resulting in reliable data and emphasizing objectivity, generalizability, and the use of statistically validated instruments [43]. This thesis employs a quantitative approach to explain the influence of various predefined factors on user activity on Mastodon instances. Additionally, a mixed-methods approach could have been employed to investigate the underlying topic. However, this dismissed given the extensive data available on Mastodon instances and the resource limitations inherent to a master's thesis.

Examples for quantitative data collection would be observation, interview, questionnaire method, the psychological test, physiological measurement and document analysis. With quantitative data collection methods, the degree of structuring is very high and fully standardised instruments are used. The informative value of a study that uses quantitative data collection methods depends on the quality of the standardised measurement instrument [28]. This thesis utilises online survey as it allows efficient, large-scale data collection directly from the target user base, making it more suitable than other methods like personal interviews or physiological measurements. A questionnaire is a targeted, systematic and rule-based generation and recording of verbal and numerical self-reports by participants on selected aspects of their experience and behaviour in written form [28]. The questionnaire will be provided via the internet, more specifically via Mastodon and the target group are active Mastodon users. The empirical study design is visualised in table 2:

Data Collection Method	CAWI (Computer Assisted Web Interview)
Population (N)	Active Mastodon users
Sampling Method	Quota sampling
Sample Size (n)	3,621
Characteristics	Monthly Activity on Mastodon
Data Collection Period	16.12.2024 - 02.01.2025
Data Collection Location	Mastodon

Table 2: Survey Design

The sampling method employed in this thesis survey is quota sampling. This method is based on the assumption that the population can be divided into groups according to specific socio-demographic characteristics that are relevant to the research question. It is a deliberate intervention in the selection process that leads to a specific composition of the sample. As a non-probabilistic sample, the resulting quota sample cannot claim global representativeness. However, it can claim characteristic-specific representativeness thanks to the quota plan [28]. The research questions addressed in this thesis are all related to user activity and the ways in which it can be improved. Consequently, it is important that the survey participants are active users of Mastodon, as this will enable the identification of the factors that are of particular importance to them in the context of their own instance. Based on the collected data the user base satisfaction on their instances should be evaluated. According to Döring [28] six main classification criteria can be used to differentiate between different written survey techniques and questionnaire forms. The following criteria will be used for this thesis:

Structuring Degree of the Survey	Standardized Questionnaire
Questionnaire Mode	Electronic Questionnaire
Questionnaire Administration	Internet-mediated Questionnaire
Type of Respondents	Affected Individuals
Survey of Individuals or Groups	Individual Survey
Special Query Format	No Special Query Format

Table 3: Questionnaire Criteria

The questionnaire will be electronic and will consist only of closed questions, thereby ensuring full standardisation. The questionnaire will be distributed via Mastodon, where it will not distinguish between mobile and desktop access, given the existence of mobile apps for certain instances. The survey will be directed towards individual Mastodon users, rather than experts or decision-makers representing other affected individuals.

8.2.2 Survey Development and Design

The construction of the standardised questionnaire is undertaken in two stages. Initially, a rough draft is created, followed by a detailed draft. Subsequently, the constructed questionnaire is subjected to a pretest, and any revisions that are necessary are made [28]. In this chapter the initial design of the survey is presented in table 4.

Measurements in quantitative social research lead to measurement values with different levels of information. A distinction is made between four levels of measurement or scales of measurement according to ascending information content. The following list includes the four different scales of measurement according to Döring (2023) [28]:

Title of the Survey	Survey on Enhancing and Sustaining User Activity on Mastodon
Questionnaire Instruction	Thank you for your interest in this study! My name is [...] and I'm a student at the University of Vienna doing research on Mastodon. My master thesis aims to understand the gap between registered accounts and active users. After a spike in registrations in 2022, the number of active users on Mastodon decreased, and this study aims to find out why. Your unique experience as an active Mastodon user is essential to this research. There are no "right" or "wrong" answers - just your honest insights. The survey is anonymous and completely voluntary, and all answers will be used for scientific purposes only. It will take about 10 minutes to complete, and your participation would be a valuable contribution to understanding the factors that influence user activity on Mastodon. If you have any questions or require further information, please email <a51807829@unet.univie.ac.at
Content-Related Question Blocks	Questions related to governance structure, technical infrastructure, and engagement. An example question might be: "The Community Guidelines on my instance are clear to me." • Fully disagree • Mainly disagree • Neutral • Mainly agree • Fully agree
Statistical Questions	Age, Country, Gender, Mastodon Instance (Optional)
Feedback	Open question for feedback: Do you have any comments about this questionnaire or this study?
End Slide	Thank you very much for your support with this study. It is greatly appreciated!

Table 4: Rough Concept of the questionnaire on user activity

- **Nominal scale**: Nominal-scale characteristics only differentiate between categories in no particular order, such as gender or marital status. The categories are mutually exclusive (e.g. smoker/ non-smoker) and cover all possible characteristics. Analysis concentrate on frequencies, whereby the numerical values can be assigned arbitrarily as long as each category is given its own number.

- **Ordinal scale**: Ordinal scales are used for variables that show an order or ranking among categories. Such categories may include company hierarchies (lower management, middle management, top management), competition placements or education levels. Ordinal scales allow to ascertain the relative position of any two categories, even if the exact numerical difference between them is not known.

- **Interval scale**: An interval scale is a quantitative method that allows for a more precise measurement of characteristics than an ordinal scale, which relies on a greater than/less than relationship. An interval scale requires the use of an empirical relative, wherein the dominance relations of all object pairs adhere to a weak order structure. In contrast to an ordinal scale, where the strength of one object's dominance over another is inconsequential, an interval scale requires that the pairwise dominance relations be ranked according to their strength.

- **Ratio scale**: Ratio-scale characteristics are defined by values that are equally spaced and, in contrast to the interval scale, also possess an absolute zero point. An illustrative example might be the number of errors a person makes in dictation or the number of children they have. In the empirical relative of a ratio scale, linking operations are typically defined in addition to a weak order relation of the objects.

This thesis used a combination of single-item and short Likert scale measures (3-5 items) to assess a single variable. A Likert scale is a psychometric scale made up of several statements that measure the same characteristic, with respondents rating their level of agreement on a scale. Typically, a 5-point scale is employed, where responses reflect varying degrees of a measured trait [28]. The questionnaire presented participants with the following options: strongly disagree, disagree, neutral, agree, and strongly agree. Critics of the orginal Likert scale argued that 10-20 nearly identical items can demotivate participants by measuring the same construct many times. In contrast, psychometric short scales, which typically use 3-5 items, are more efficient. The most effective approach is to select the best and most selective items from a longer Likert scale, thereby allowing for a more focused and engaging measurement [28].

8.2.3 Computational Study

In addition to the survey, a computational study on Mastodon was conducted. This comprised a collection of digital data, including highly informative behavioural traces, referred to as digital trace data. These digital traces may reflect various aspects of users' lives, including social relationships, group affiliations, political views, consumption patterns and more. They are non-reactive, generated naturally in everyday life, rather than created in response to the researchers' actions [28]. In computational research, data quality is similar to that in observational studies, where users are unaware that their data are being analysed for research purposes. This lack of user awareness raises important ethical and legal considerations [65]. Therefore, when conducting a computational study, it is important to consider the following four ethical principles by Matthew Salganik [71]:

- Respect for Persons involves treating individuals as autonomous agents and honouring their wishes. This principle suggests that researchers must obtain participants' consent and letting them, not the researchers, decide.
- Beneficence is about analysing the risks and benefits of a study and then deciding, whether they strike an appropriate ethical balance.
- Justice means that researchers should not be allowed to prey on the powerless and vulnerable, and that the benefits and risks of a study should be distributed fairly.
- Respect for Law and Public Interest explicitly urges researchers to take a broader view and to include law in their considerations. One of these aspects would be compliance, where researchers should seek to identify and comply with relevant laws, contracts and terms and terms of service. Another would be transparency-based accountability, which implies that researchers should be clear about their aims, methods and results at all stages and take responsibility for their actions.

In their 2019 study, Possler et al. identify six methods of collecting research data in computational studies. These include cooperation with companies that own data, purchasing data from data owners, application programming interfaces (APIs), data scraping, tracking, and secondary analysis of existing data sets [65]. This thesis will make use of APIs and therefore this topic will be explained in more detail. API is an application programming interface that allows connection to a digital platform and access to the data in a defined and often strictly limited way. Some social media platforms offer open or public APIs so that researchers can access the platform and data (e.g. Reddit, Wikipedia) [28].

However, other platforms do not provide open APIs or the free APIs have different features and limitations. There may be limits on the time span and amount of data that can be retrieved via the API, and often there is no transparency about the procedures used to generate data output for API requests [65]. An API can be accessed for data collection or extraction using appropriate tools. This can be done either with self-programmed tools (e.g. created with programming languages such as Python or Java), with ready-made tools (e.g. with corresponding packages of the R software), or with user-friendly web services that allow the API to be called via a simple input mask on the web [28]. Data analysis in research projects should follow a structured process: first, data is imported from sources such as files or APIs. Next, tidying organises it into a consistent format, while transformation refines it by adding new variables, filtering or merging. Visualisation then reveals patterns, and modelling applies statistical techniques to uncover insights or make predictions. Finally, clear communication, often through reports or presentations, shares the results effectively [86]. For a better understanding, all of the steps are depicted in figure 5.

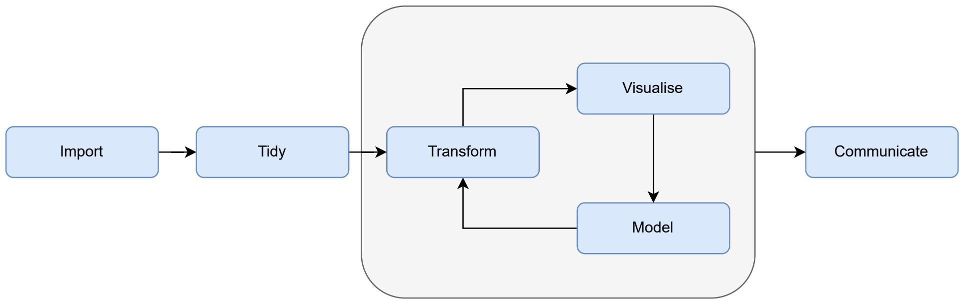

Figure 5: Data Processing Cycle in Computational Studies (based on Wickham et al., 2019) [86]

For this study, Python was used to retrieve data from two different open Mastodon API sources. As Mastodon is an open source platform, there aren't as many restrictions as with the centralised social media platforms mentioned above, and there are also providers offering free APIs. After calling on the API, the retrieved data will be stored in a structured format within a DataFrame. Once stored, the data will undergo a thorough pre-processing phase which includes filtering, cleaning, and formatting to ensure accuracy and consistency. This process transforms the raw API data into a form suitable for detailed analysis, highlighting relevant factors such as user activity and its influencing factors such as instance size, transparency, or latency. Finally, the processed data was analysed to identify patterns and correlations, and then visualised to illustrate key findings and insights. The resource requirements for API data collection are relatively low, and the results are often machine-readable, well structured, and rich in metadata [65].

This thesis employs a multi-method approach, comprising a quantitative questionnaire and a computational study. In previous studies, this method has been as extensive as linking social media data with survey data to combine the distinctive strengths of each data collection method [15]. However, linked data can be disclosed and sensitive, and in this case, the computational data was collected from instances and not from single users, whereas the online survey would require the opposite approach. Therefore, it was decided that the data would be researched in parallel, rather than linked. The objective was to determine which factors correlate with user activity on Mastodon. The survey elicits users' perceptions of the importance of specific factors, while the computational study analyses digital trace data from Mastodon instances. For example, governance structures such as the transparency of community guidelines or the presence of proactive moderators were included in the online survey, as these are more subjective perceptions for participants. Conversely, latency or uptime can be measured and tracked, and therefore they were evaluated in the computational study.

8.3 Operationalisation of Variables

This section presents an analysis of the variables employed in this thesis and their operationalisation. In quantitative research, variables are distinguished according to their nature: continuous variables possess an infinite number of infinitely finely graded characteristics, whereas discrete or discontinuous variables can only assume a small and limited number of characteristics. Operationalisation involves selecting observable variables and measurement instruments to determine the characteristics of a theoretical concept. It often uses multiple indicators rather than a single one. This process provides a concrete operational definition for a theoretical concept [28].

8.3.1 Dependent Variable

In order to explore the influence on the dependent variable, it is typically necessary to record the values of the dependent variables in as fine a gradation as possible [28]. In this thesis, the operationalisation of the dependent variable is the frequency of user activity on Mastodon, as in the number of times a user accesses the platform within a specific period of time. An active user was considered as an individual who logs into Mastodon at least once a month. While the computational study was limited to the examination of monthly active users, the questionnaire permitted the differentiation of active users based on their frequency of use: daily, multiples times per week, weekly, or monthly. This allowed a more detailed examination of the activity levels.

8.3.2 Independent Variables

While dependent variables are often continuous characteristics that are measured with as many gradations as possible, independent variables (i.e. the causal factors under consideration) are usually discrete variables with few characteristics [28]. Table 5 presents an overview of the independent variables and provides a description of each:

Independent Variables	**Description**
Instance Size	Measured by the number of registered users on a Mastodon instance.
Inbound Engagement	The amount of boosts, replies and favourites a user receives on their toot
Outbound Engagement	User interactions with other users' content, measured by the total number of likes, comments, and shares.
Active Contribution	The frequency of a user tooting content.
Community Guidelines	The transparency, clarity, and inclusivity of community guidelines on an instance.
Content Moderation Practices	The extent of enforcement of community guidelines by moderators, including removing harmful content, issuing warnings, and blocking violating instances.
Information Overload	The volume and complexity of information users are exposed to in their feeds.
Instance Uptime	The availability of a Mastodon instance, measured by uptime.
Security Measures	The extent of security protocols implemented.
Instance Transparency	Existence of terms of service and privacy policies on an instance.
Sustainability	The extent to which an instance is 'green' or environmentally sustainable.
Latency	The server response time, measured by the delay users experience when interacting with the instance.

Table 5: Overview of Independent Variables and their Measurement Description

9 Results of Computational Analysis

9.1 Data Collection and Preprocessing

The principal application programming interface (API) tool employed in this thesis was Fediverse Observe [2] which permitted the querying of numerous independent variables specified in the previous chapter. First, the data were retrieved from the API with the following GraphQL query:

Listing 1: GraphQL query for retrieving data

```
{
  nodes {
    id
    active_users_monthly
    {param_field}
    total_users
  }
}
```

For the activity rate, data was needed on both the number of monthly active users and the total number of registered users. The dynamic parameter field was modified in accordance with the specific variable in question. After the retrieval of the data, the total number of entries was found to be 21,074 instances. However, since an initial assessment of the data indicated outliers in the user counts, it was necessary to filter and clean the raw data. It was decided to use a lower and upper percentile of 0.1 percent to eliminate unrealistically high numbers of registered users. Moreover, an instance was classified as active only if it had at least one active user, meaning instances with no active users were excluded from the analysis. To ensure reliability of the results, single-user instances and instances with a higher number of active users than registered users were excluded. Single-user instances, by definition, have an activity rate of 100% (since both the active and registered user counts are identical). Additionally, smaller instances were removed to prevent the skewing effect of unusually high activity rates, which could distort the overall patterns observed in larger, more active communities. In the end, the data frame for each independent variable was saved in a CSV file to ensure the reproducibility of the results.

9.2 Sample Overview and Descriptive Statistics for the Dependent Variable

This section provides an overview of the sampled instances and a descriptive analysis of the activity rate, defined as the ratio of monthly active users to total registered users. Previous research has suggested that engagement levels vary depending on instance size. For example, the moderator of Mastodon.at estimated that an optimal community size for engagement falls between 500 and 1,000 users [94]. However, to establish a more data-driven threshold, a computational analysis was conducted, testing various size classifications. The results indicated that a threshold of 100 users effectively captured behavioural differences between smaller and larger instances. Consequently, instances were categorised as small (2–100 users) and large (more than 100 users). This division allows for a more precise examination of user activity patterns and engagement trends across differently sized communities. The total number of instances after filtering was 11,038, representing 8,603,613 registered users and 761,236 monthly active users. This figure comes close to the official count of approximately 9 million users reported by Mastodon Analytics in October 2024 [49].

Table 6 illustrates that the analysis includes a wide range of instance sizes, from 1 to the maximum of 98 for small instances and 101 to 615,171 registered users for large instances. The mean number of registered users per instance for small instances was 11.31, with a standard deviation of 17.05. For large instances, the mean was 5,001.77, with a standard deviation of 25,633.42. Upon dividing the sample into smaller and larger instances, it becomes evident that the mean activity rate is higher in the smaller instances than in the

[2] https://api.Fediverse.observer

larger ones respectively 51.18% and 16.83%. The minimum activity rate of large instances with 0.005% is notably low. However, it should be noted that the activity rate column was subjected to a filtering process to remove outliers, corresponding to the 0.1% and 99.9% quantiles of the data. This ensured that extremely low activity rates were removed.

Descriptive Statistic	Small Instances	Large Instances
Activity Rate		
Mean	51.18%	16.83%
Std. Dev.	28.58%	15.91%
Min	1.64%	0.005%
Max	100%	99.61%
Total Users		
Mean	11.31	5,001.77
Std. Dev.	17.05	25,633.42
Min	2	101
Max	98	615,171
Monthly Active Users		
Mean	4.09	425.59
Std. Dev.	6.49	1,844.60
Min	1	1
Max	94.00	44,632.00
Overall statistics		
Instances	9,339	1,699
Total Users	105,601	8,498,012
Monthly Active Users	38,163	723,073

Table 6: Descriptive Statistics for Small versus Large Instances

9.3 Independent Variables

This chapter presents the results of the correlation analysis between the independent variables and the user activity rate. There are two correlation coefficients utilized in this thesis, which will be explained shortly. The Pearson coefficient is used to assess linear relationships in data that is normally distributed. Whereas the Spearman coefficient is used to evaluate monotonic associations in non-normal, ordinal, or outlier-prone datasets. Both coefficients range from -1 to +1, with 0 representing no correlation [75]. Table 7 shows the correlations between the independent variables and user activity.

Variables	Pearson (Small)	Spearman (Small)	Pearson (Large)	Spearman (Large)
Latency	-0.0035 (ns)	-0.0023 (ns)	0.0165 (ns)	0.0161 (ns)
Uptime	0.0213 (s)	0.1078 (s)	0.1003 (s)	0.2448 (s)
Security	-0.0102 (ns)	-0.0103 (ns)	-0.0660 (s)	-0.0663 (s)
Transparency	-0.173 (s)	-0.160 (s)	0.116 (s)	0.167 (s)
Sustainability	-0.019 (ns)	-0.0184 (ns)	0.029 (ns)	0.030 (ns)
Instance Size	-0.137 (s)	-0.636* (s)	-0.137 (s)	-0.636* (s)

Table 7: Results of the Computational Analysis - Pearson and Spearman Correlation Coefficients for Small and Large Instances. Statistically significant correlations are marked with "s" and strong correlations with an asterisk (*).

9.3.1 Latency

This subchapter presents the results of the correlation analysis between latency and user activity rate. In this thesis, latency refers to the time required to connect to a Mastodon server, as measured by Fediverse Observer. This metric, recorded by the platform, was later retrieved via an API call. Before conducting the correlation analysis, the dataset for was explored and visualised. The preliminary analysis showed several outliers that were subsequently removed based on 1% quantiles. This left 9,238 small instances and 1,677 large instances for the correlation analysis.

The descriptive statistics for latency indicated a mean of 0.15 with a standard deviation of 0.13 for large instances and a mean of 0.16 with a standard deviation of 0.14 for small instances. This already demonstrates that the overall latency on Mastodon instances is relatively low with larger instances showing less latency than smaller instances. The Pearson correlation coefficient for small instances was -0.0035 with a p-value of 0.7374 and for large instances, it was 0.0165 with a p-value of 0.4988. The Spearman correlation coefficient for small instances was -0.0023 with a p-value of 0.8244, and for large instances, it was 0.0161 with a p-value of 0.5090. Hence, there is no statistically significant correlation between latency and user activity rate. Figures 6 and 7 present the findings of the latency analysis for small and large instances, respectively.

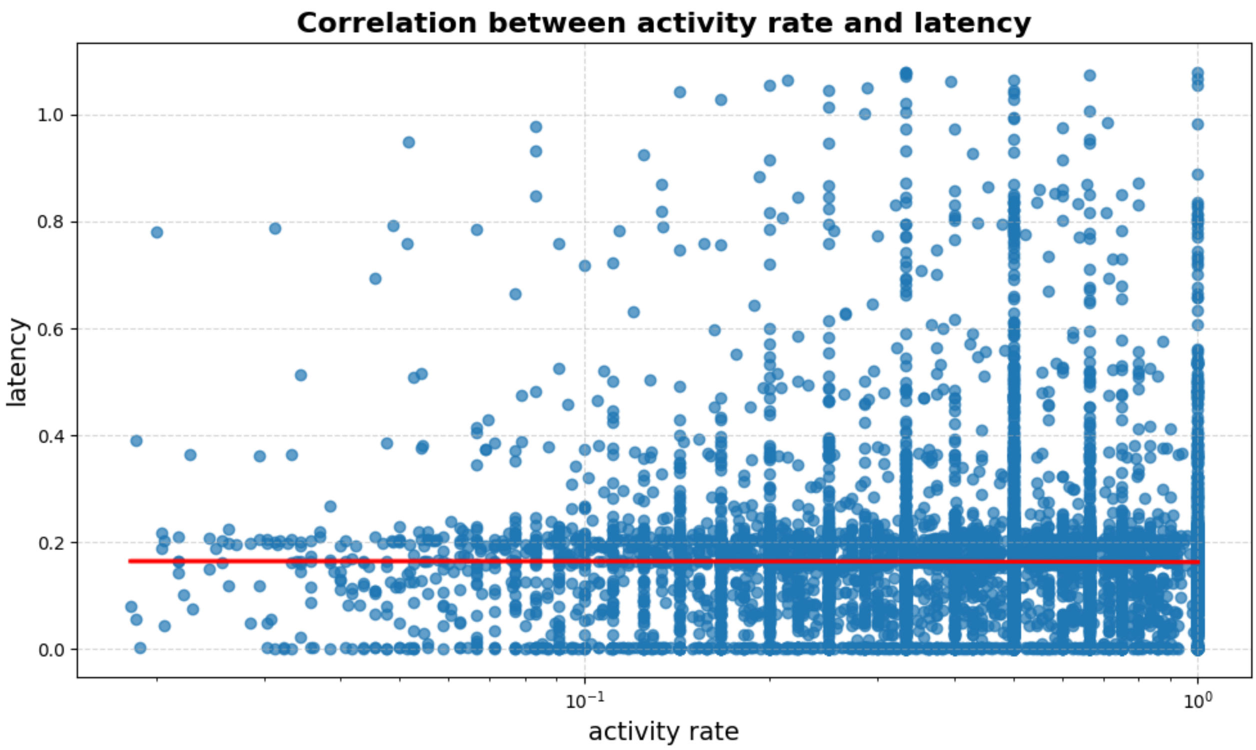

Figure 6: Relationship between latency and activity rate on small instances (<100 registered users). Each data point represents one instance, with the regression line shown in red.

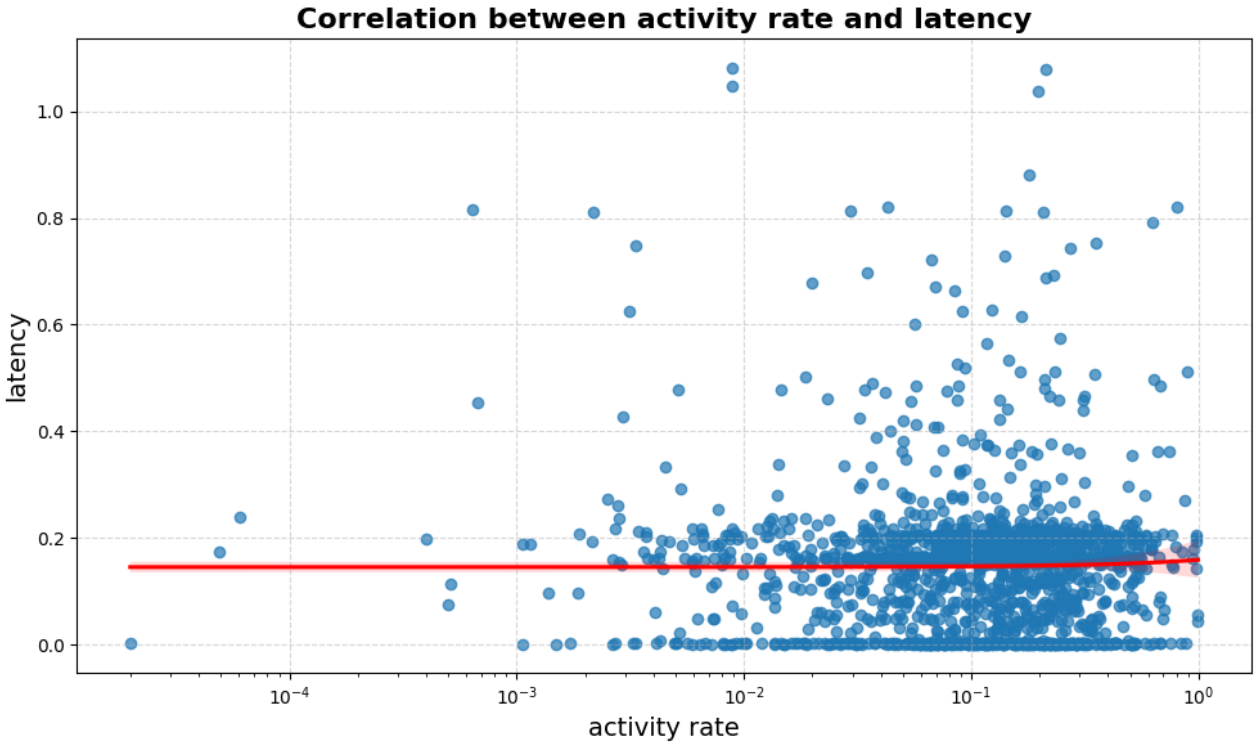

Figure 7: Relationship between latency and activity rate on large instances (>=100 registered users). Each data point represents one instance, with the regression line shown in red.

9.3.2 Uptime

This chapter presents the results of the correlation analysis between uptime and user activity. Uptime refers to the time (in percentage) a server is available to users. The descriptive statistics for small instances indicate a mean of 98.45%, a standard deviation of 3.58, with values ranging from a 45.8% to 100% uptime over 9,339 instances. Large instances recorded an average uptime of 99.05%, with a standard deviation of 1.85, a minimum of 78.56%, and a maximum of 100% uptime from a total of 1695 instances. Like the latency variable, every instance indicates a relatively high uptime, with larger instances exhibiting greater uptime compared to smaller ones.

Small instances revealed a Pearson correlation coefficient of 0.0213 and a p-value of 0.0392, suggesting a weak positive correlation that is statistically significant. The Spearman correlation coefficient was 0.1078 ($p<0.001$), suggesting a substantial rank-order association. For larger instances, the Pearson correlation coefficient was 0.1003 ($p<0.001$), showing a weak but statistically significant linear correlation between uptime and activity rate. The Spearman correlation coefficient was 0.2448 ($p<0.05$), indicating a stronger rank-order correlation. Although instances with higher uptime tended to have higher user activity, the correlation remains weak. Figures 8 and 9 illustrate the relationship between uptime and user activity for both smaller and larger instances.

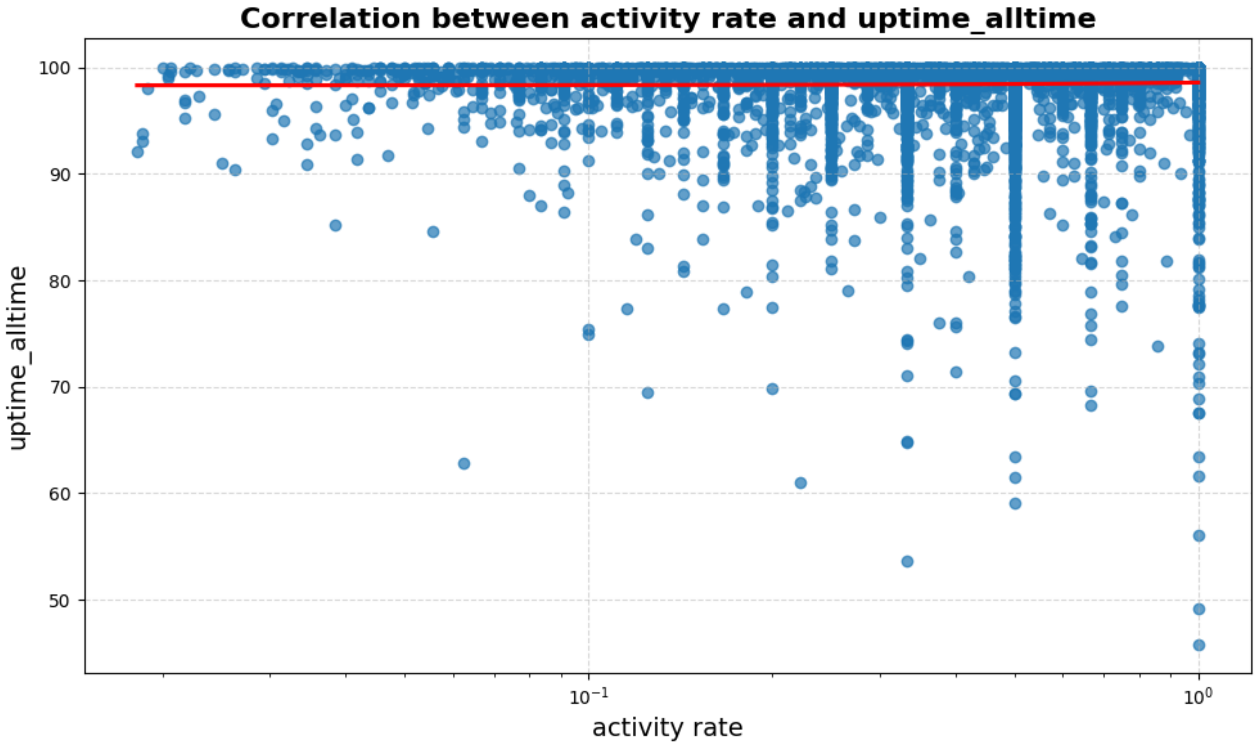

Figure 8: Relationship between uptime and activity rate on small instances (<100 registered users). Each data point represents one instance, with the regression line shown in red.

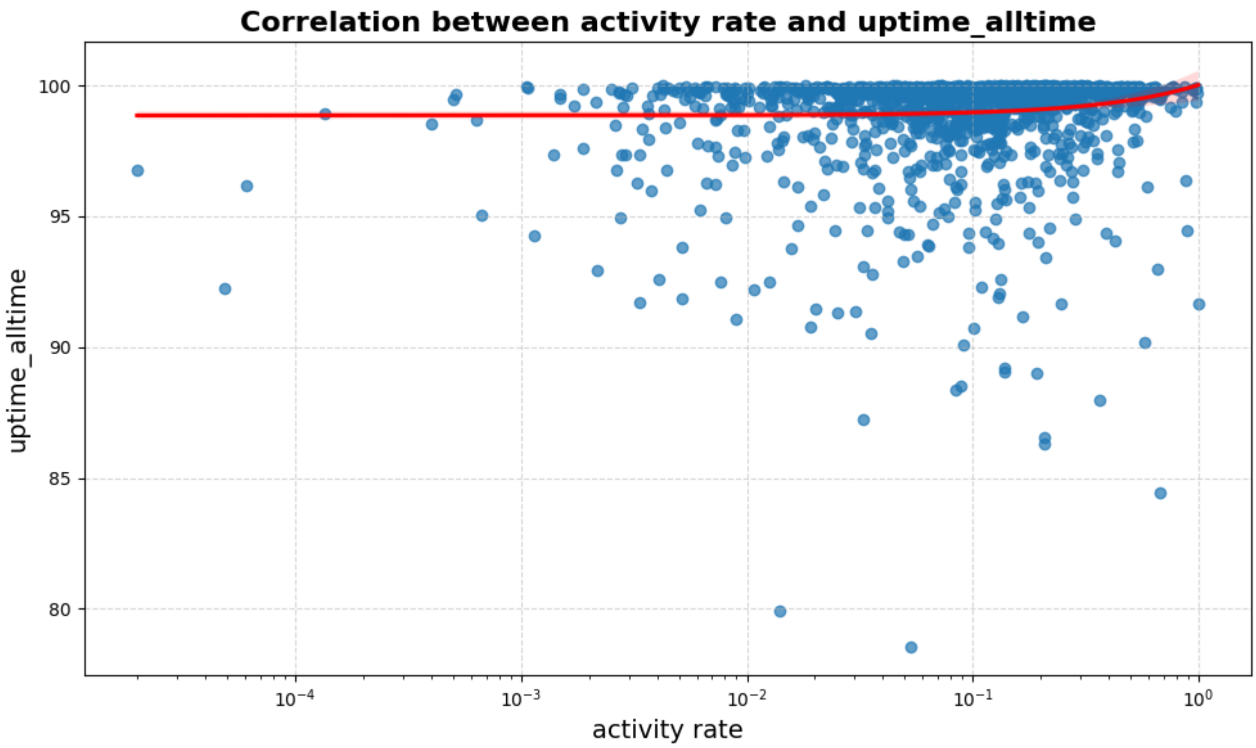

Figure 9: Relationship between uptime and activity rate on large instances (>=100 registered users). Each data point represents one instance, with the regression line shown in red.

9.3.3 Security

The security measures in the computational analysis were based on two key measurements: whether an instance had DNSSEC (Domain Name System Security Extensions) enabled and whether it had a valid SSL (Secure Sockets Layer) certificate. However, an examination of the descriptive statistics of SSL implementations revealed that 99.97% of instances exhibited valid SSL certificates. Therefore, the SSL implementations were excluded from the analysis, given the lack of variability in the values. In contrast, DNSSEC was implemented on only 19% of small instances and 21% of large instances, making it a superior candidate.

The analysis of small instances revealed a weak correlation between DNSSEC implementations and activity rate. The Pearson correlation coefficient was -0.0102 ($p = 0.3254$) and the Spearman correlation coefficient was -0.0103 ($p=0.3215$), indicating that no statistically significant relationship exists. Conversely, the analysis of large instances yielded statistically significant results. The Pearson correlation coefficient was -0.0660 ($p<0.05$) and the the Spearman correlation was -0.0663 ($p<0.05$). While DNSSEC implementations exhibited a negative correlation with activity rate, its impact is relatively weak. The distributions of data points for small and large instances are illustrated in Figures 10 and 11, respectively.

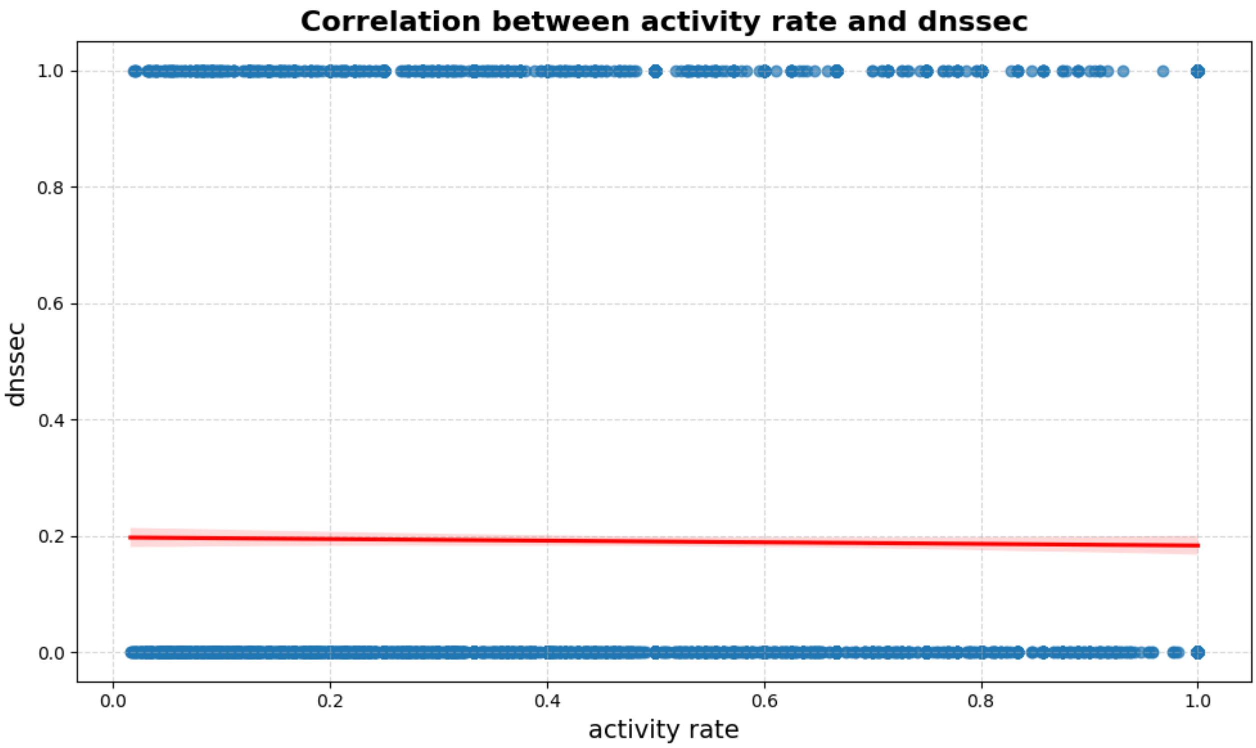

Figure 10: Relationship between DNSSEC implementation and activity rate on small instances (<100 registered users). DNSSEC implementation is represented as a binary variable (1 = implemented, 0 = not implemented). Each data point represents one instance, with the regression line shown in red.

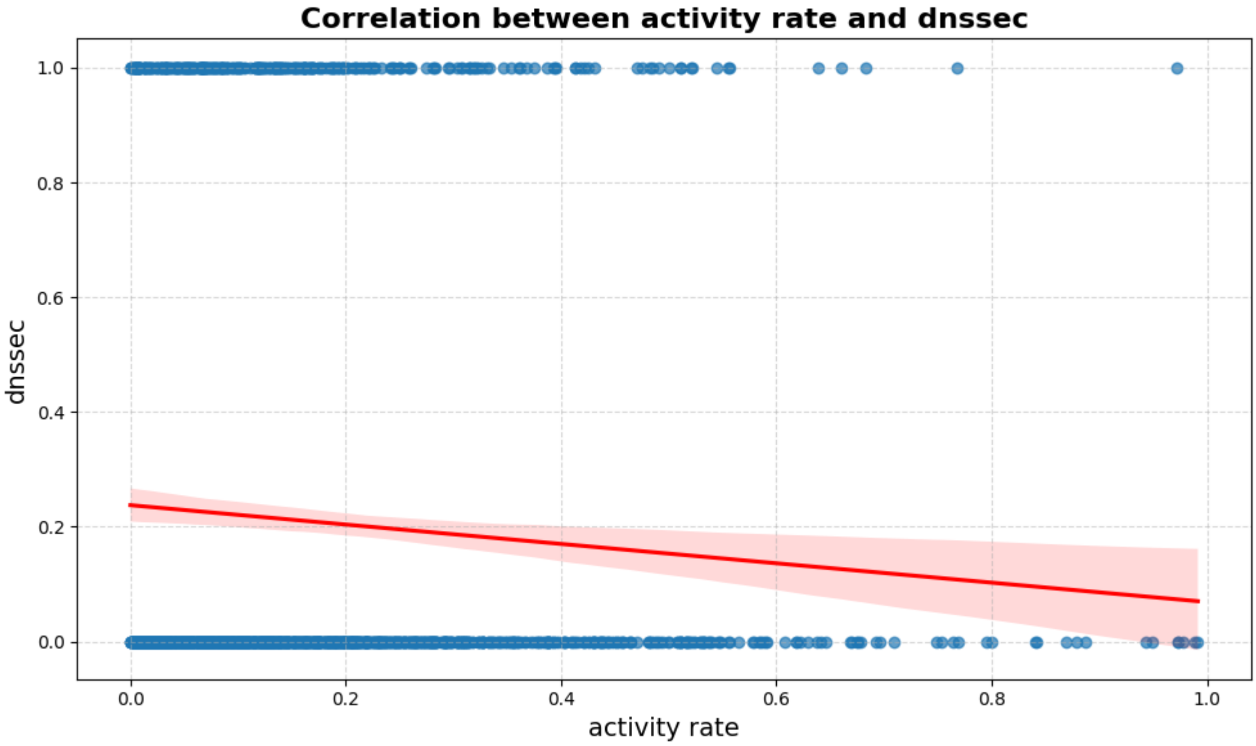

Figure 11: Relationship between DNSSEC implementation and activity rate on large instances (>=100 registered users). DNSSEC implementation is represented as a binary variable (1 = implemented, 0 = not implemented). Each data point represents one instance, with the regression line shown in red.

9.3.4 Transparency

This chapter presents the findings of the analysis between transparency and user activity rate. Transparency was measured by two components - privacy policy and terms of service and whether an instance had them linked in their profile. The descriptive statistics demonstrated that among the 9,334 small instances, 80.51% had a privacy policy, and 79.82% had terms of service linked. Among the 1,693 large instances, 87.30% had privacy policies, and 86.24% had terms of service linked. This demonstrates that larger instances are more likely to have privacy policies and terms of service linked.

Small instances, had a negative correlation between linking terms of service and activity rate, with a Pearson correlation of -0.166 ($p<0.001$) and a Spearman correlation of -0.152 ($p<0.001$). Similarly, privacy policies and user activity rate on small instances, showed a weak negative correlation with a Pearson correlation of -0.180 ($p<0.001$) and a Spearman correlation of -0.167 ($p<0.001$). This shows a modest, negative impact of transparency on activity rates, with an average Pearson coefficient of -0.173 and an average Spearman coefficient of -0.160.

Conversely, the relationship between transparency and user activity rates demonstrated a positive and statistically significant correlation for instances of a larger scale. The correlation between linking terms of service had a Pearson coefficient of 0.105 ($p<0.001$) and a Spearman coefficient of 0.152 ($p<0.001$). The relationship between privacy policies and activity rates was slightly stronger, with a Pearson correlation of 0.126 ($p<0.001$) and a Spearman correlation of 0.182 ($p<0.001$). The mean of the privacy policies and terms of service yielded a Pearson coefficient of 0.116 and a Spearman coefficient of 0.167. These results indicate that on larger instances greater transparency is valued. The correlation between privacy policies and user activity for smaller and larger instances are illustrated in Figures 12 and 13 and the correlation between terms of service and user activity are depicted in Figures 14 and 15.

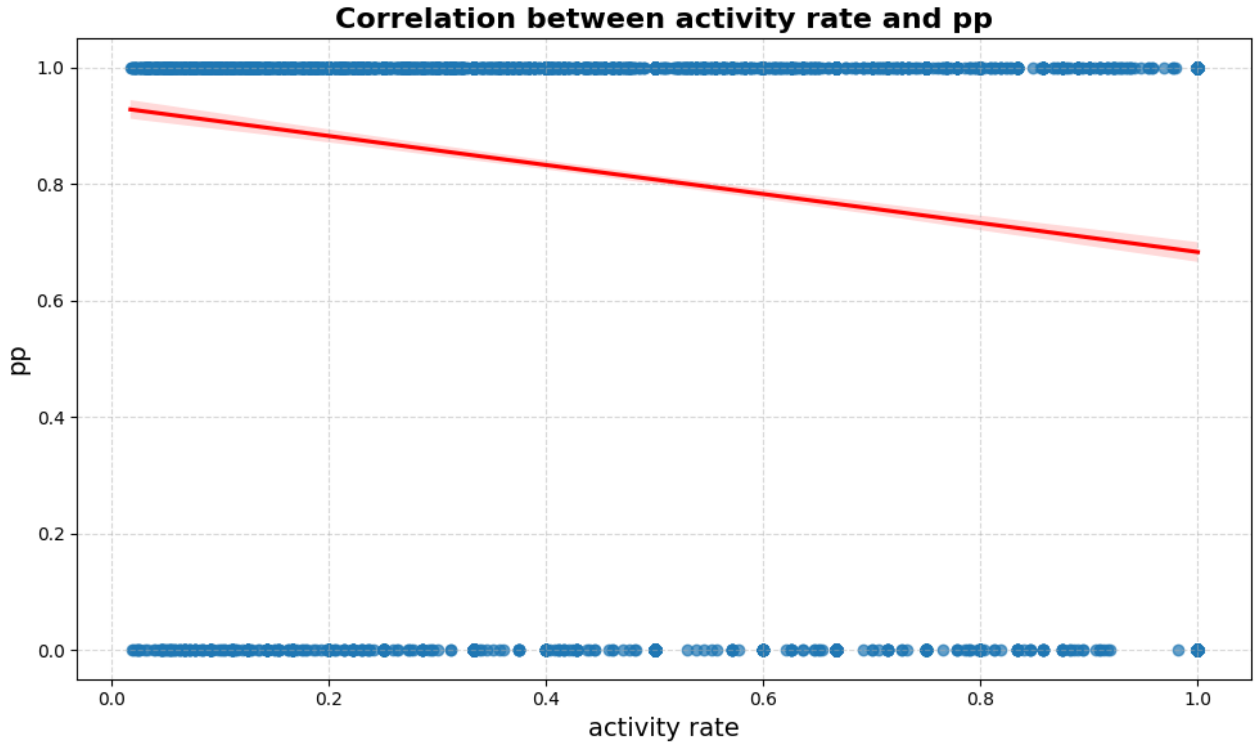

Figure 12: Relationship between linked privacy policies on an instance and activity rate on small instances (<100 registered users). The presence of a linked privacy policy is represented as a binary variable (1 = linked, 0 = not linked). Each data point represents one instance, with the regression line shown in red.

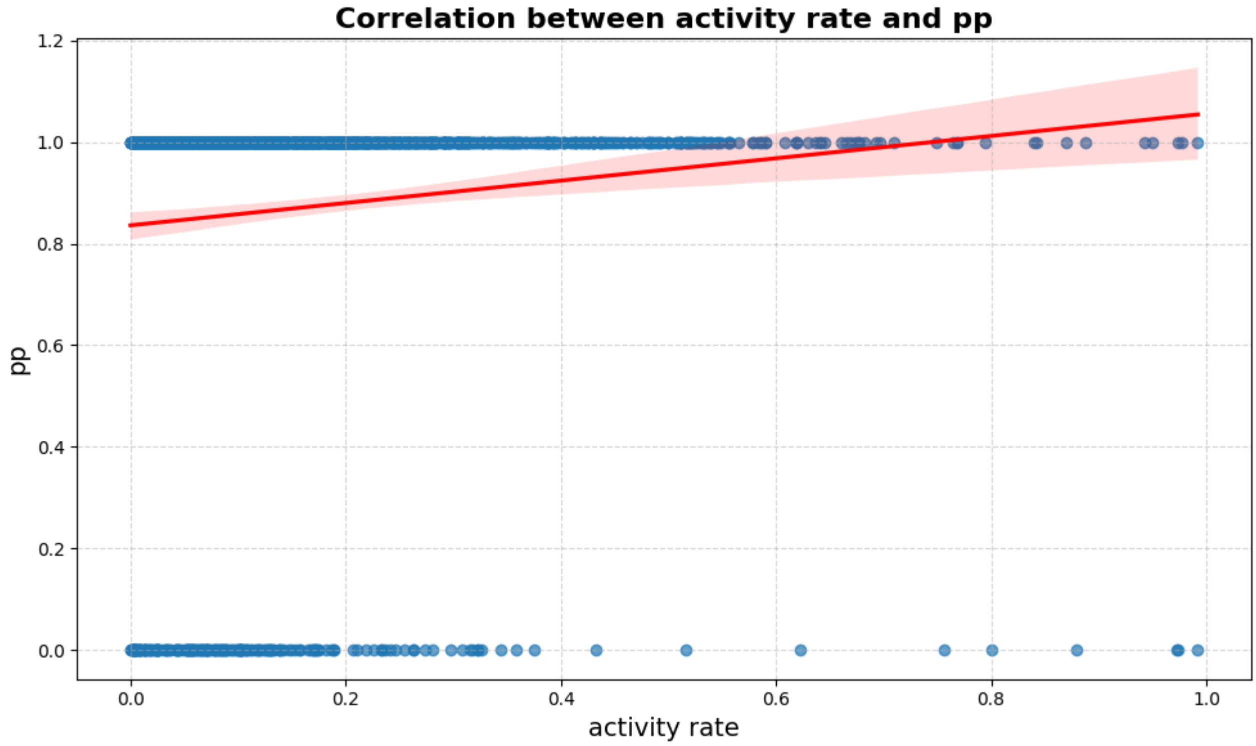

Figure 13: Relationship between linked privacy policies on an instance and activity rate on large instances (>=100 registered users). The presence of a linked privacy policy is represented as a binary variable (1 = linked, 0 = not linked). Each data point represents one instance, with the regression line shown in red.

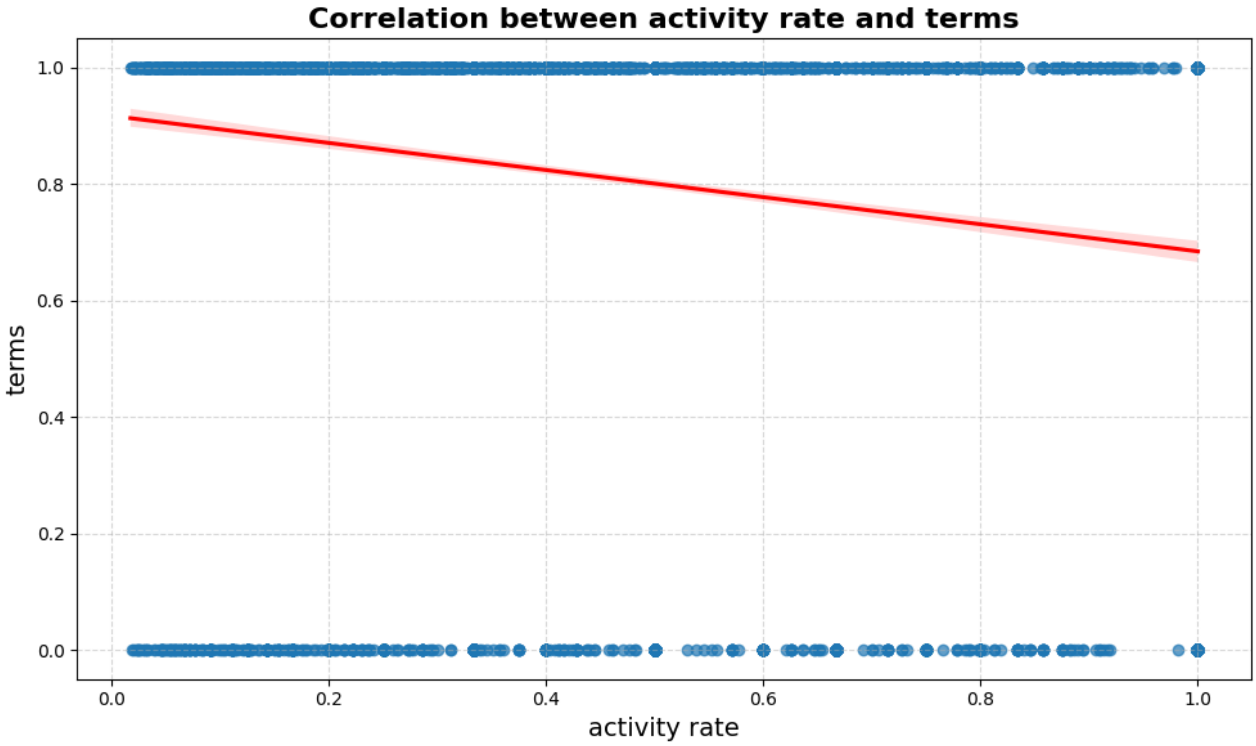

Figure 14: Relationship between linked terms of service on an instance and activity rate on small instances (<100 registered users). The presence of linked terms of service is represented as a binary variable (1 = linked, 0 = not linked). Each data point represents one instance, with the regression line shown in red.

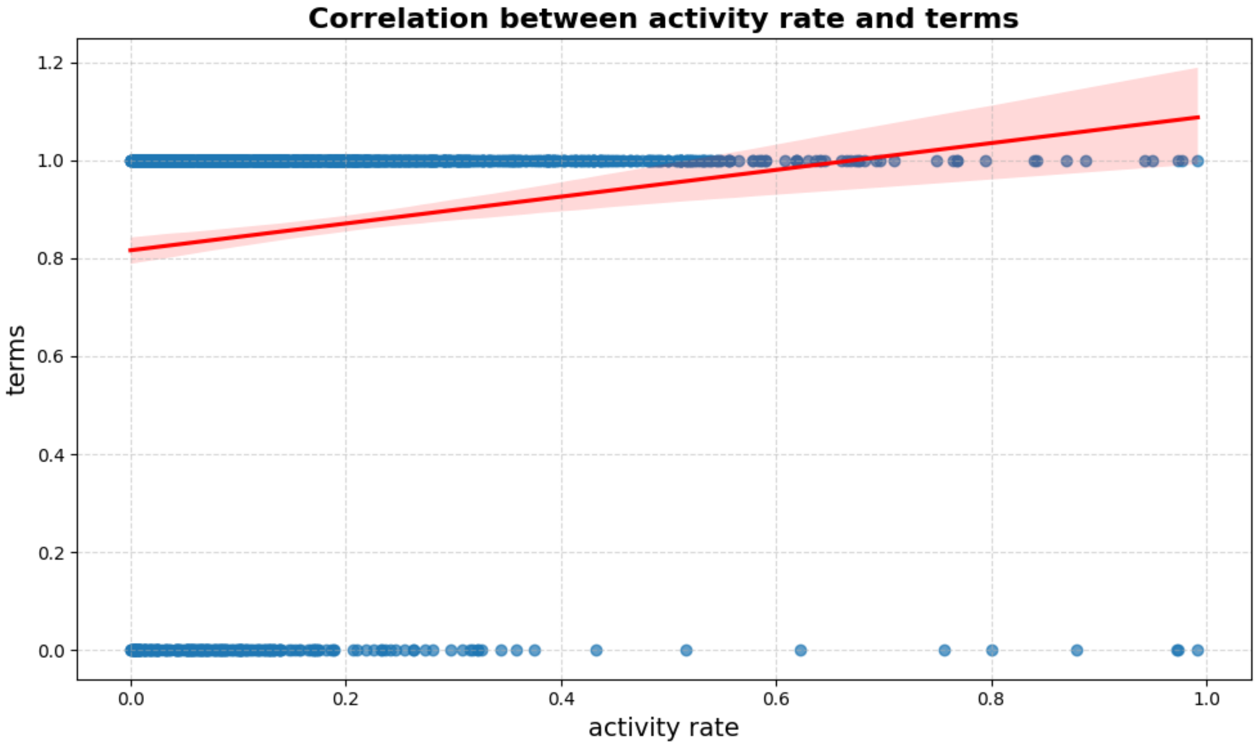

Figure 15: Relationship between linked terms of service on an instance and activity rate on large instances (>=100 registered users). The presence of linked terms of service is represented as a binary variable (1 = linked, 0 = not linked). Each data point represents one instance, with the regression line shown in red.

9.3.5 Sustainability

The sustainability in the computational analysis was measured by a Boolean value that indicates whether a Mastodon server is hosted at a green data centre. Descriptive statistics for small instances indicated that approximately half of the servers are green hosts (50,97%), while a greater proportion of large instances (58,99%) are hosted on a green data centre. The analysis of small instances demonstrated a negative, weak correlation that is not statistically relevant. The Pearson correlation revealed a coefficient of -0.019 (p = 0.069) and the Spearman correlation a coefficient of -0.0184 (p = 0.075).

The correlation analysis of large instances revealed a Pearson coefficient of 0.029 (p-value = 0.346), and a Spearman coefficient of 0.030 (p= 0.219). The visualisation of the results for small and large instances are illustrated in Figures 16 and 17.

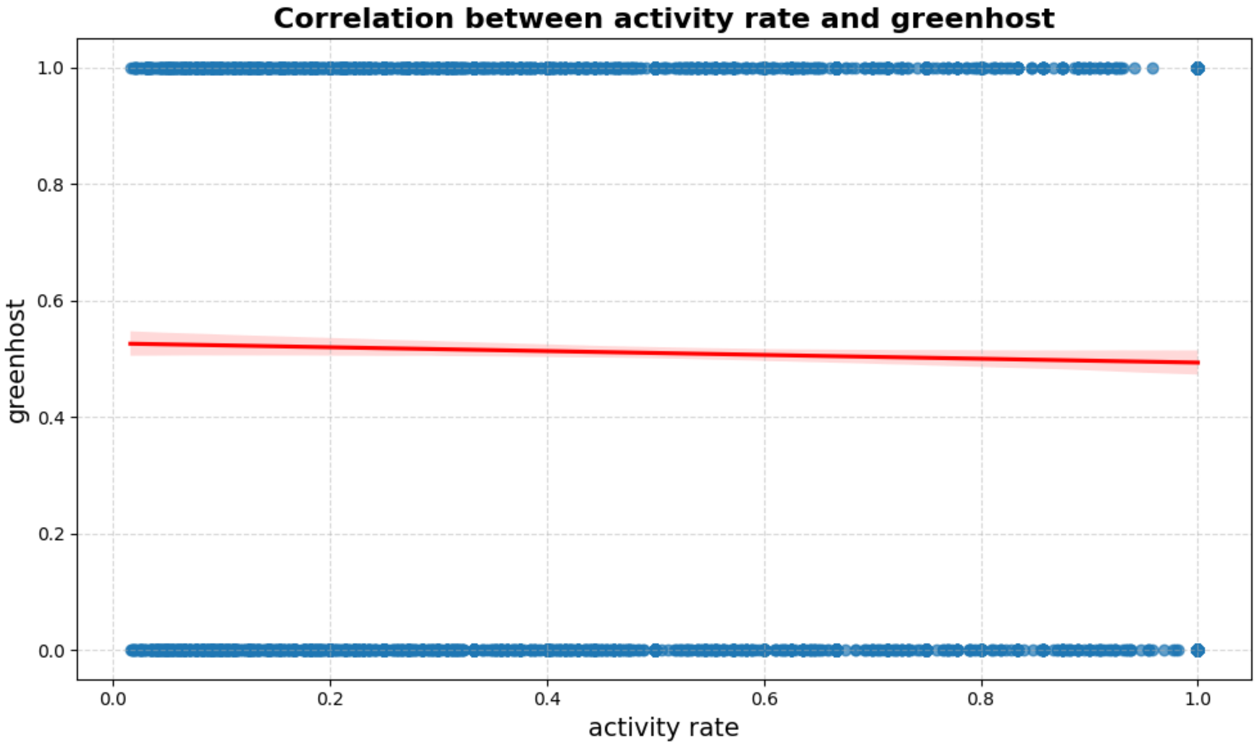

Figure 16: Relationship between an instance that is hosted at a green data centre and activity rate on small instances (<100 registered users). The "green instances" are represented as a binary variable (1 = hosted at a green data centre, 0 = not hosted at a green data centre). Each data point represents one instance, with the regression line shown in red.

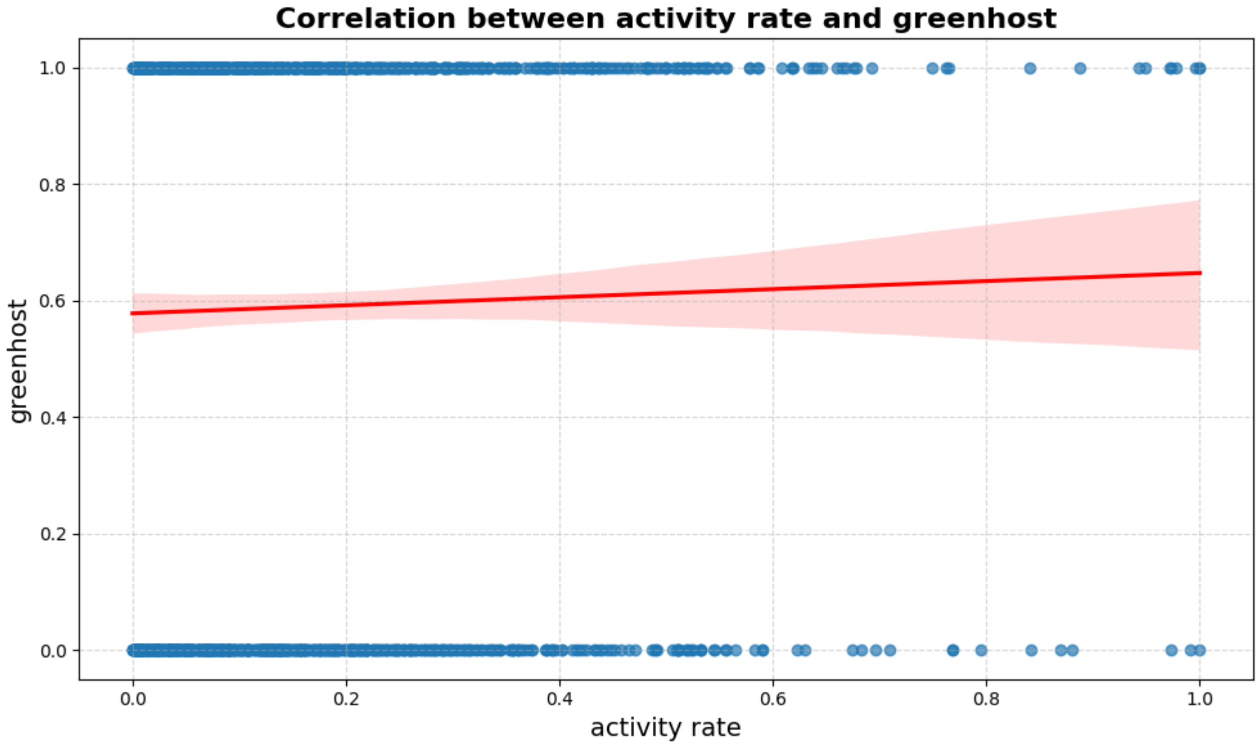

Figure 17: Relationship between an instance that is hosted at a green data centre and activity rate on large instances (>=100 registered users). The "green instances" are represented as a binary variable (1 = hosted at a green data centre, 0 = not hosted at a green data centre). Each data point represents one instance, with the regression line shown in red.

9.3.6 Instance size

This chapter presents the findings of the correlation analysis between instance size and user activity. Instance size was the sole independent variable not divided into larger and smaller instances, as it is a factor affecting activity rates similarly across all instances. This approach enabled a uniform analysis without categorising instances by size. The descriptive statistics showed an average of 525.39 users with a standard deviation of 4,480.22, reflecting significant variability in instance sizes. The median (50th percentile) number of registered users was 6, indicating that half of the instances had fewer than 6 users. The lower quartile (25th percentile) had only 3 users, while the upper quartile (75th percentile) reached 28 users, showing that most data rows were small instances. However, the maximum number of registered users in a single instance reached 179,064, which is smaller than the largest instance on the Mastodon network, due to the elimination of outliers in the previous steps. These statistics show the differences in instance sizes, with a few large instances significantly skewing the overall distribution. The analysis revealed a statistically significant inverse correlation between the number of registered users on an instance and the user activity rate. The Pearson correlation for total registered users revealed a coefficient of -0.137 ($p<0.001$) and the Spearman correlation revealed a coefficient of -0.636 ($p<0.001$). Figure 18 demonstrates the correlation between registered users and user activity.

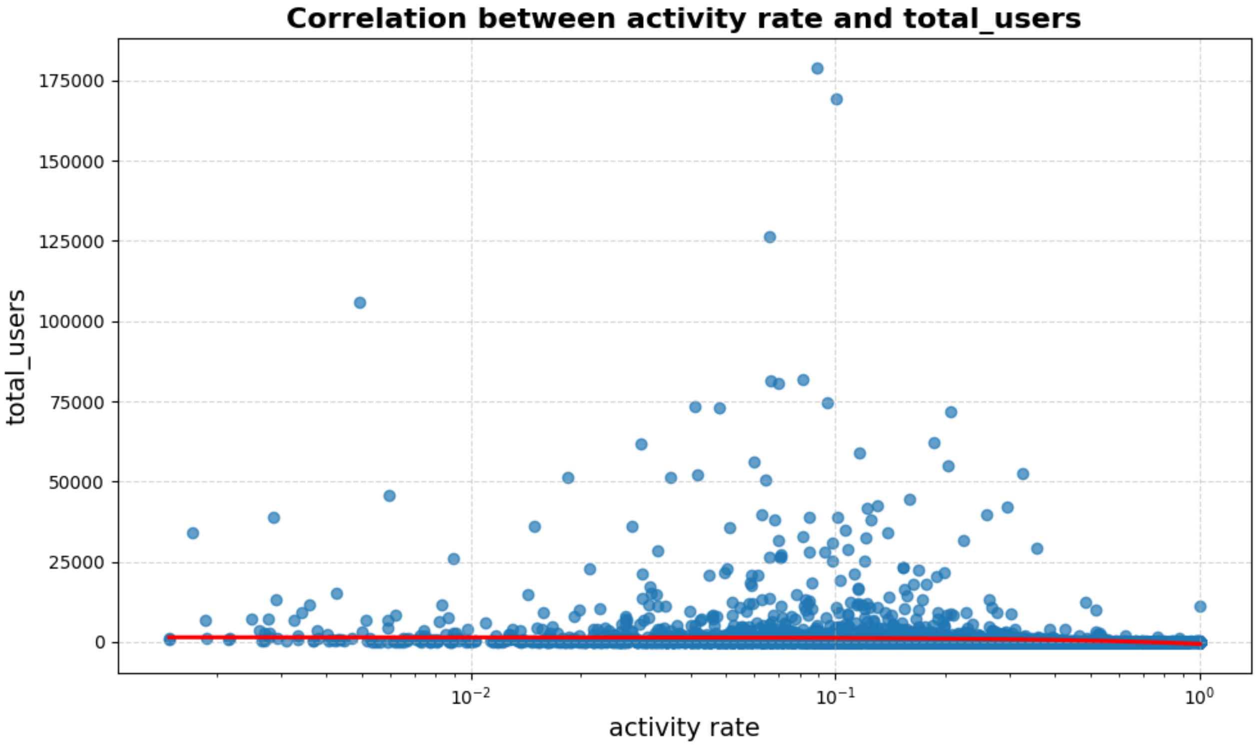

Figure 18: Relationship between instance size and user activity rate. Each data point represents one instance, with the regression line shown in red.

10 Results of the User Survey

10.1 Data Collection

This chapter describes how the data for the questionnaire was collected. Before the survey was published, a pre-test with seven Mastodon experts was conducted. The tool used to create the survey, SoSci Survey [3], offered the possibility to provide feedback directly on the questionnaire with input fields that could later be viewed directly by the author in the tool's interface. This made providing and incorporating feedback very efficient. Over the course of a month, the pre-test was online and feedback was continuously provided and incorporated into the final draft. The main adjustments were made to clarify the wording of some questions. For example, in the original questionnaire, the options for user activity were less than once a week and a few times a month. The less than once a week option was removed due to unclarity reasons. Adjustments were also made to make the options clearer and easier to understand, such as replacing 'engagement' with 'interaction', or rephrasing the explanation of 'sustainability' in the technical factors section. Another very important issue raised was the inclusion of optional answers in each question for the statistical summary. This meant that if participants did not feel comfortable answering a question, there was always an "I prefer not to say" option.

After incorporating all the adjustments from the pre-tests, the final survey was published on Mastodon on 16/12/2024. The survey targeted monthly active Mastodon users and was consequently distributed on the Fediverse, given Mastodon's affiliation with the broader Fediverse. Consequently, the post was also visible to users of all other Fediverse platforms. This has led to dissatisfaction among some Fediverse users, who were unable to participate in the survey. This issue will be addressed further in the limitations section. On the first day of publication, only one user completed the survey. The strategy was therefore adapted to ask people with a wider following to help distribute the survey. A number of Mastodon users agreed to share the survey, resulting in a continued accumulation of responses. However, after a few hours one of the posts received a backlash in the form of negative replies the distributor decided to delete the survey from their timeline. Concurrently, some users became aware of the original post and started boosting it, leading to a rapid distribution of the survey. Over the subsequent two days, more than 2,000 people completed the questionnaire. In total, the original post on Mastodon was shared 1.3 thousand times, resulting in 11,783 clicks to the survey and 4,070 responses. Of these, 3,621 participants completed the survey, while 449 dropped out, resulting in a dropout rate of 11%. Approximately 100 participants exited the survey after the consent question about the data privacy policy, which was presented at the beginning of the questionnaire. An additional 157 participants dropped out after the third question, which asked about the social media platforms they use. This question did not only serve as an icebreaker, but also as a filter to ensure that only Mastodon users proceeded to complete the rest of the survey.

The survey was constantly monitored, and a key benefit of utilising Mastodon for its distribution was the ability to receive feedback in real-time. An email address was also provided in the survey, which resulted in a constant flow of questions and feedback via email. During the survey, it was discovered that the design was not optimised for administrators and moderators, with some respondents expressing uncertainty regarding the questionnaire's instructions. This issue was addressed promptly by incorporating an additional filter question for these specific cases. Following these adjustments, only a few minor technical issues were encountered, and the survey received predominantly positive feedback. In Figure 19, the response statistics are illustrated, with the orange bars representing the finalised questionnaires and the grey segments indicating the total number of interviews over the entire interview period.

[3] https://www.soscisurvey.de/en/index

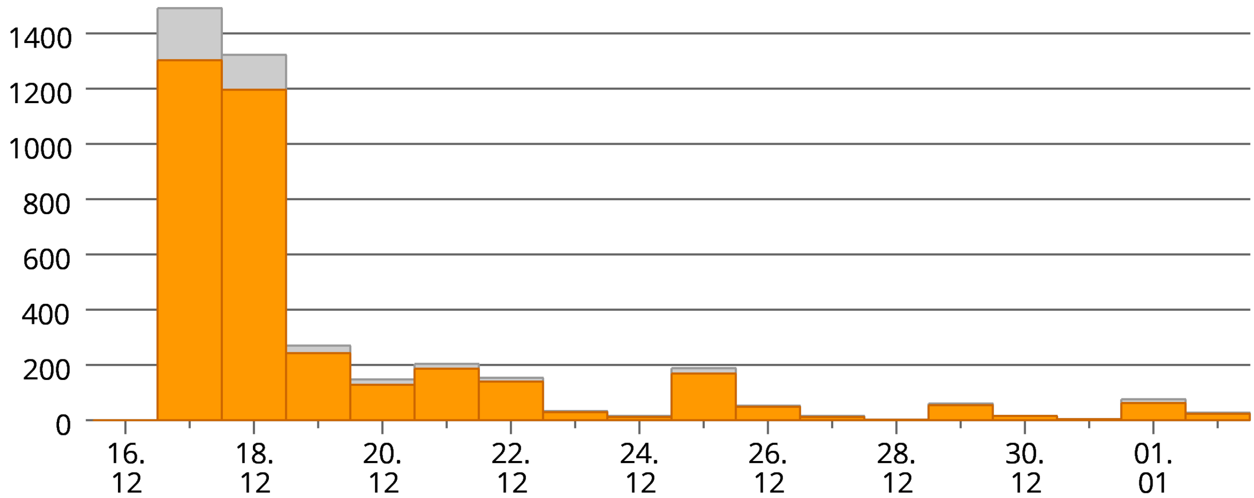

Figure 19: Response statistics of the user survey conducted from 16 December 2024 to 2 January 2025

10.2 Data Processing

Following the data collection, a CSV file containing the survey data and the corresponding codebook were downloaded from the SoSci Survey Tool. The settings for downloading had already permitted some preliminary processing, as only interviews that had been completed were downloaded, i.e. interviews in which the participants had reached the final page of the questionnaire. This reduced the amount of interviews from 4070 to 3621, including the pretests data. The CSV file and the codebook were subsequently imported into Python using the pandas library. The codebook comprised the variable abbreviation, the variable description, the response code and label, as well as the variable and the input type. The most relevant information was then transformed into a JSON dictionary. After importing the data, missing values were identified and replaced with a placeholder value to facilitate further analysis. The following Python code was used to achieve this:

```
raw_data = raw_data.fillna("Missing")
```

Additionally, the SoSci Survey tool recorded the time spent overall on the survey, with an average duration of 350 seconds, or approximately five minutes. Assuming that participants completing the questionnaire in significantly less time would provide unreliable responses and potentially distorting the results, it was determined that data entries with a recorded completion time of less than one minute should be excluded from the analysis.

It is crucial to ensure the data meets the appropriate requirements: either jointly normally distributed for Pearson correlation or ordinal/non-normally distributed continuous data for Spearman correlation [75]. During the review of the dictionary, it was observed that the user activity and several other factors were arranged in the wrong order. This was due to adjustments made to the response options during the creation of the survey. Therefore, the data required rearranging as the correct sequence plays a significant role in the correlation analysis. Furthermore, the daily usage of Mastodon was combined with the monthly usage. More specifically, the first question posed to participants enquired about the frequency of Mastodon usage, with respondents given a range of options from 'a few times a month' to 'every day.' Participants who selected 'every day' were prompted to specify the amount of daily usage. In the preprocessing step, the two responses were combined. For instance, '5 minutes' from the second question was adjusted to '5 minutes daily', while the 'every day' option from the first question was removed.

This thesis employed a combination of single-item measures and short Likert scales (3 to 5 items) to evaluate individual variables. The Likert scale is a psychometric tool consisting of multiple items designed to assess the same variable. Respondents indicate their level of agreement on a 5-point scale in order to capture different intensities of the measured trait [28]. In the survey of this thesis, participants had the following response options: strongly disagree, disagree, neutral, agree, and strongly agree.

Since the study employed a Likert scale, ensuring data reliability during evaluation was important and for that Cronbach's alpha reliability coefficient was used. Cronbach's alpha is a statistical measure used to evaluate the reliability of a set of measurements, typically when these measurements represent different occasions, raters, alternative forms, or, most commonly, questionnaire or test items. When applied to questionnaire items, it evaluates internal consistency, describing the extent to which the items in a scale measure the same underlying construct. A higher Cronbach's alpha value suggests greater internal consistency among the items, meaning they are likely assessing the intended concept reliably [12]. In this study, Cronbach's alpha is used to determine whether the Likert scale items consistently capture the same characteristic, ensuring that the responses accurately reflect the measured construct.

The value range of the reliability coefficient is from 0 (completely unreliable measurement, consisting only

Category	**Cronbach's Alpha**
Community Guidelines	0.83
Engagement on My Posts	0.68
Engagement General	0.78
Information Overload	0.81
Technical Factors	0.61
Moderation Practices	0.65
Transparency	0.74

Table 8: Cronbach's alpha reliability scores for measured independent variables.

of random measurement errors) to 1 (perfectly reliable measurement, completely unaffected by measurement errors) [28]. Reliability coefficient above .90 are generally considered to be high, and those above .80 are considered adequate [17]. However, other papers indicate that a coefficient from 0.70 is acceptable with a strong indication, that authors should interpret the value of alpha according to what they are seeking to measure and the number of items included in the scale [82, 81]. In order to achieve a satisfactory level of internal reliability, it was decided to adjust the items associated with the variable 'engagement on my posts'. Three questions were posed in this category:

1. When I toot something, I usually get some engagement on it.
2. I am satisfied with the level of engagement I receive on my toots.
3. I wish there would be more engagement on my toots (reversed).

Following the removal of question one, an increase in the Cronbach's alpha coefficient was observed, indicating a discrepancy in satisfaction levels and the fact of how much engagement is received on the post. Consequently, the decision was made to remove question one from the engagement group. Additionally, the internal reliability of technical factors was found to be comparatively low. Therefore, it was decided to measure these factors as single items, as was initially designed. Finally, the last group "moderation practices" also had to be modified to achieve a satisfactory Cronbach's alpha. The initial 4 questions were:

1. I often come across content in my feed that I think should be removed by moderators. (Reversed)
2. I feel like moderators are taking action to enforce the community guidelines on my instance.

3. I feel like moderators on my instance reply quickly to reports.

4. I believe moderators on my instance should take more action. (Reversed)

The group of questions contained two reversed items, namely questions 1 and 4. The Cronbach's Alpha increased in the absence of these items. Consequently, questions 2 and 3 will be included in the group to enhance internal reliability. The updated categories with the corresponding Cronbach's Alpha are illustrated in table 9.

Category	Cronbach's Alpha
Community Guidelines	0.83
Engagement on my Posts	0.74
Engagement General	0.78
Information Overload	0.81
Moderation Practices	0.81
Transparency	0.74

Table 9: Updated Cronbach's alpha reliability scores for measured independent variables.

After the calculation of Cronbach's Alpha, the subsequent step was to indexed the categories mentioned above, calculate their mean and round to two decimal places. The computed value was then added as a new column to the data frame. Below is an example of how one of the five categories was processed:

```
if all(col in data.columns for col in cg_columns):
    data['CG_Composite_Avg'] = data[cg_columns].mean(axis=1).round(2)
```

The final stage of the preprocessing stage was to save both the dictionary and the preprocessed dataframe. After that, the preprocessing stage was concluded.

10.3 Sample Overview

This chapter presents an analysis of the characteristics of this thesis' survey participants. It begins by outlining the demographic profile of the respondents, followed by a detailed examination of their activity levels on Mastodon. While 3,621 users completed the survey, the statistics presented reflect only those participants who answered the specific question. Respondents who skipped questions or selected "prefer not to answer" are excluded from the analysis and graphs.

Firstly, the distribution of countries is presented, with Figure 20 demonstrating that a significant proportion of survey participants are from the United States and Germany, which is in alignment with Figure 2. Figure 2 presented the monthly active users by country measured by the instance location from the computational analysis. However, Figure 2 demonstrates that Germany has a higher number of active users, whereas in the survey 657 participants came from the United States, 344 from Germany, and 292 from Cambodia. Cambodia is an outlier in this regard, and it was not depicted in the original data frame of the computational analysis. The Netherlands, Canada, France, Australia, Spain, Finland and Austria were also depicted among the top 15 countries using Mastodon in Figure 2. Therefore, some similarities can be observed.

During the processing of the countries, certain obstacles had to be overcome. A minor inaccuracy in the survey design, wherein the term "Wohnort" (place of residence) was erroneously employed to denote country distribution, resulted in at least 367 German-speaking participants entering their place of residence in the open text field rather than their country name. This issue was not confined to German-speaking participants; it was also observed among other participants which subsequently required manual resolution through a normalization map.

This map assigned various user inputs, such as city names or alternate country names, to a standardized, canonical country name, ensuring consistent data for analysis. For example, variations like "Toulouse" or "France (but roaming around)" were normalized to "France", and "Untied Snakes" or "Banana Republic of USA" were normalized to "United States". This manual normalisation helped to reconcile the different ways in which users might refer to the same country and it also allowed for the recognition of regions that were not explicitly included in the original list, such as Catalonia, which some people considered to be their place of residence rather than Spain.

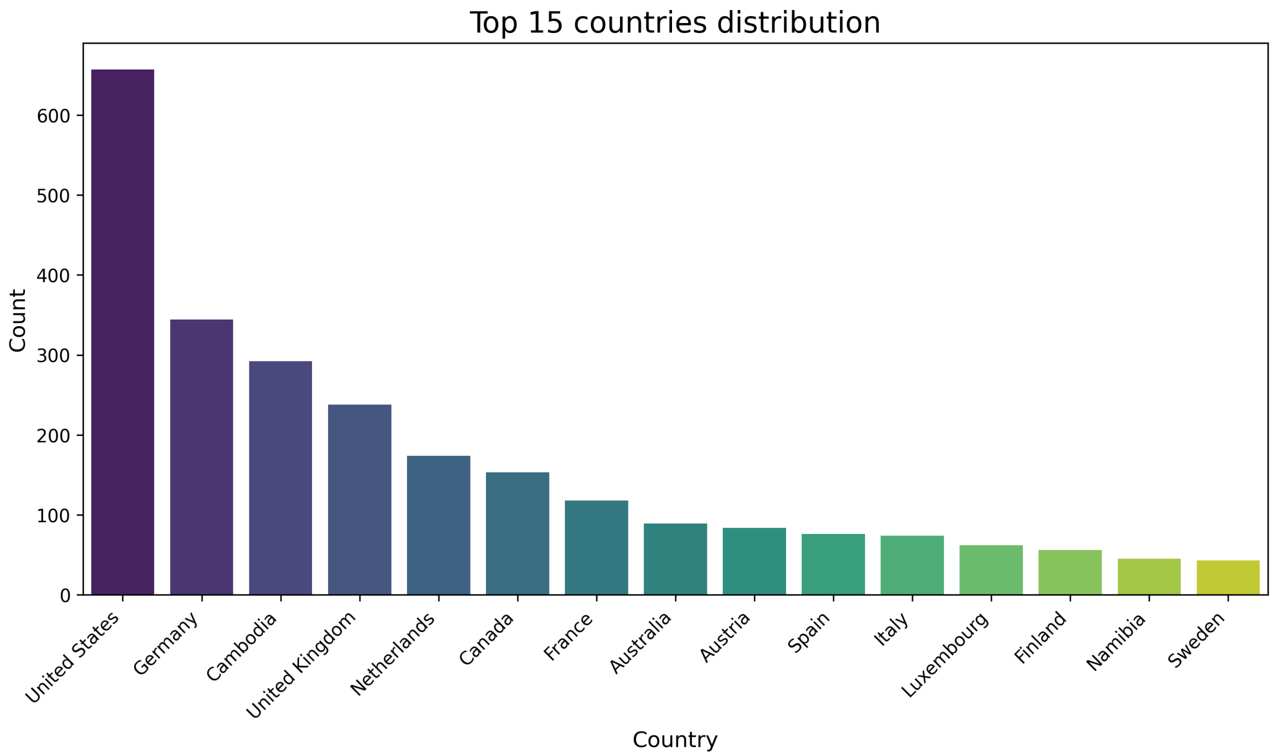

Figure 20: Distribution of participants of the Mastodon user survey by country

As illustrated in Figure 21, the country map demonstrates the global distribution of participants, with different colours representing the density of respondents from each country. The United States and certain European countries demonstrate higher participation rates in comparison to other nations. Grey areas on the map indicate countries with no recorded participants in the dataset. Nevertheless, it is evident that a significant number of countries worldwide contributed to the survey. In total, participants from 90 different countries were represented. From these participants a total of 2,331 users completed the survey in English, while 1,271 participants completed it in German.

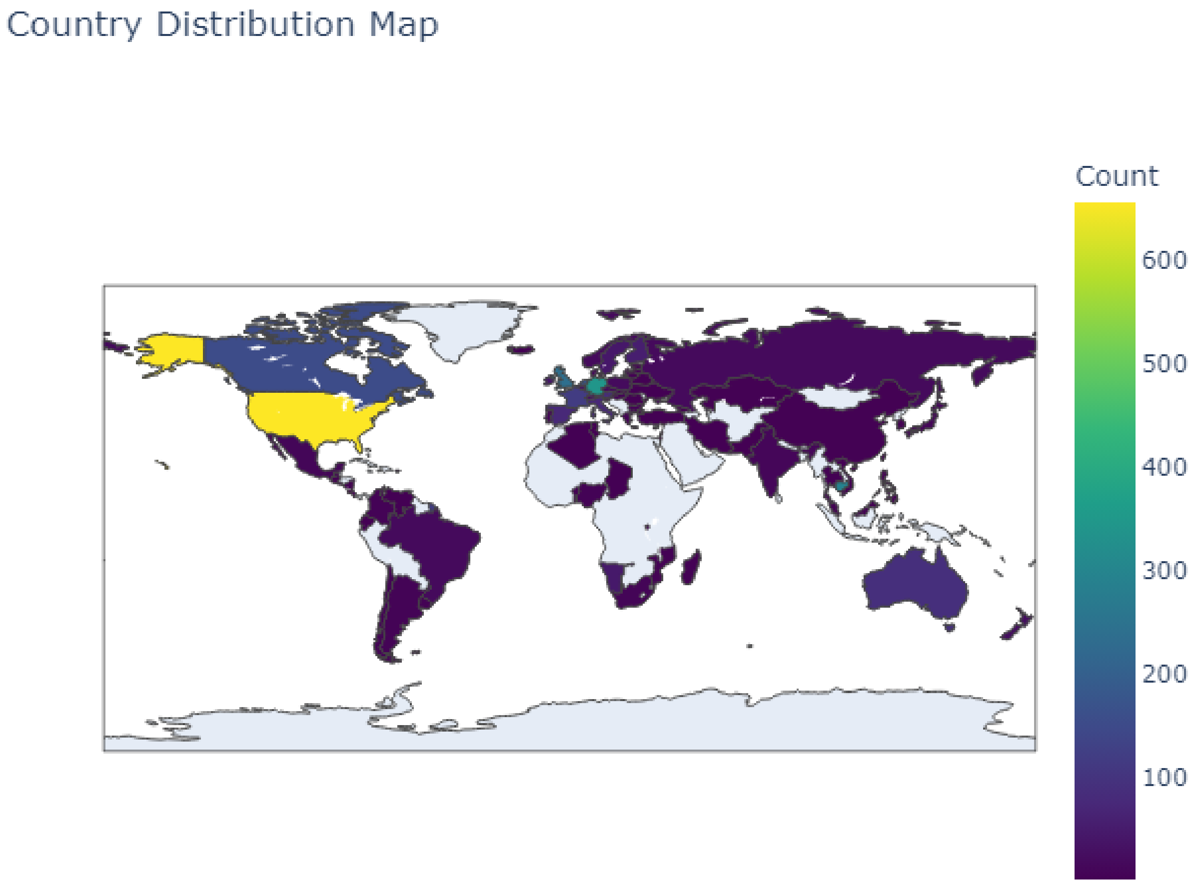

Figure 21: World map showing the distribution of user survey participants by country

The gender distribution among respondents is presented in Figure 21: 2,455 participants (74%) that replied to the gender question identified as male, 613 participants (18%) identified as female, and 216 participants (6%) identified as non-binary. Furthermore, 53 participants did not align with any of the predefined categories and instead provided their own gender specifications. These responses were subsequently grouped and represented in Figure 23.

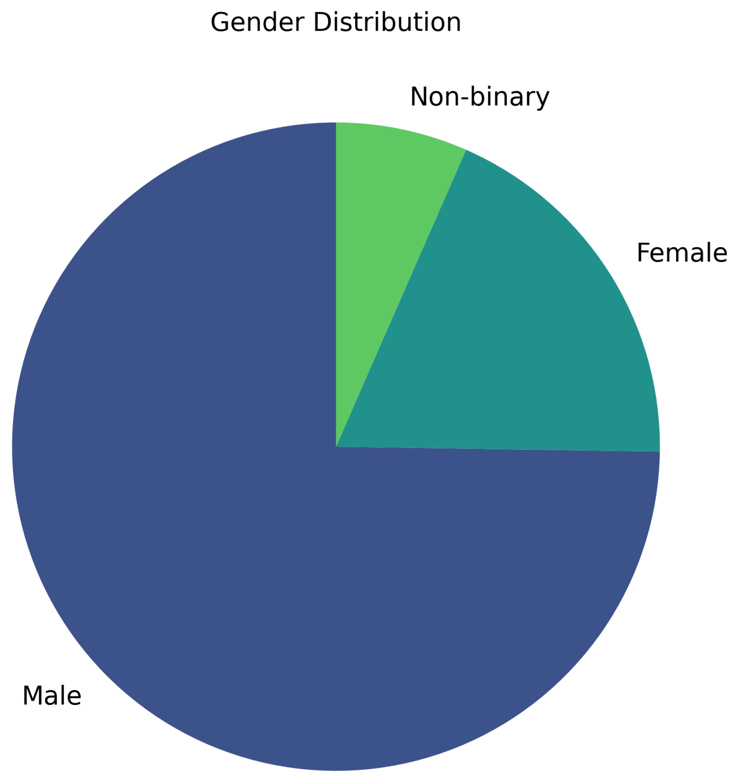

Figure 22: Gender distribution of user survey participants

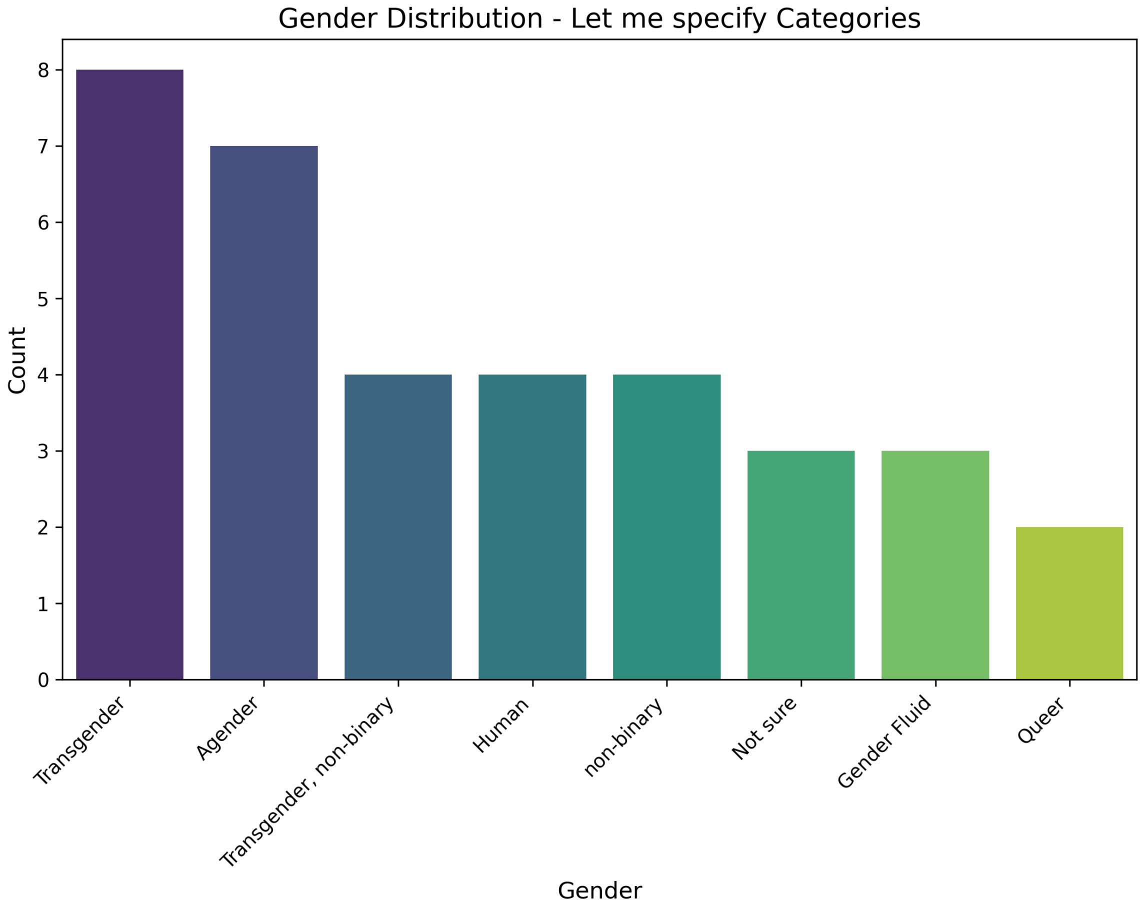

Figure 23: Detailed gender distribution of user survey participants (self-specified responses)

Prior to the analysis of the age distribution, it was necessary to preprocess the data due to the open-text format of the age responses. To address this, the ages were grouped into specific categories, and only responses within the valid age range of 12 to 109 years were considered for further analysis. The largest group were individuals aged 45-54 (28.1%), followed by 35-44 year old individuals (26.1%). The proportion of younger participants was comparatively low, with those under the age of 18 having the smallest representation (0.4%), followed by the 18-24 age group with 4.9%. The 25-34 age group represented 14.4%, those aged 55-64 17.6% and the age group 65 and over 8.5%. The detailed distribution is illustrated in figure 24, demonstrating a wide age range, with the majority of participants falling within the middle-aged categories.

10.4 Dependent Variable

In order to understand the dynamics of user behaviour on Mastodon, it is important to explore the concept of user activity. User activity is defined as the amount of time a user is active on Mastodon. The survey was designed to obtain specific amounts of time, rather than only the monthly activity as described in the computational analysis. Consequently, two obligatory questions were dedicated to the amount of time the participants spent on Mastodon. The first question related to the amount of time users spent per month on Mastodon. If participants selected the 'Never' option, the survey ended. Conversely, if the 'Every day' option was selected, participants were asked an additional question about their daily activity. Most of the survey participants indicated that they are on Mastodon every day. In total, 2952 participants selected "Every day," 500 selected "Several times a week," 49 chose "Few times a month," 40 selected "Once a week," and 14 participants responded with "Never." Figure 25 illustrates the monthly user activity, where the largest partition of 'every day' can be clearly distinguished from the other options.

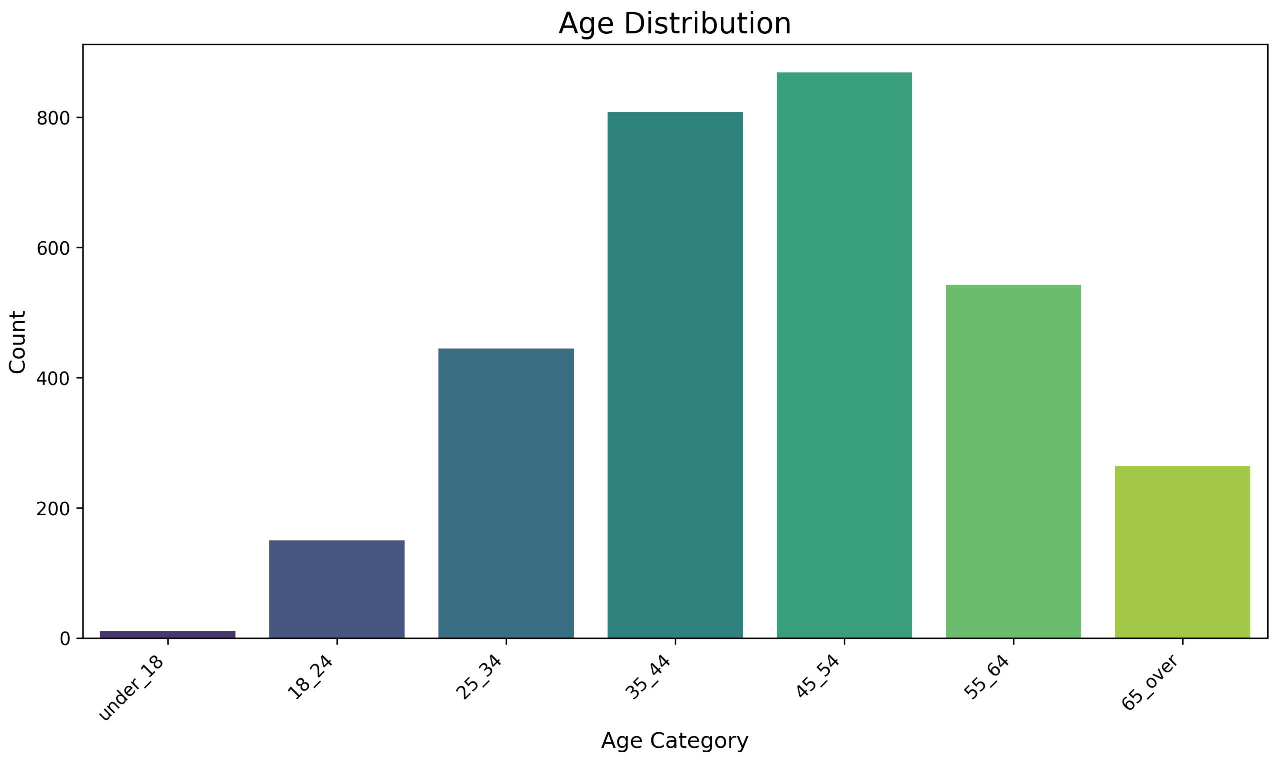

Figure 24: Distribution of user survey participants by age group

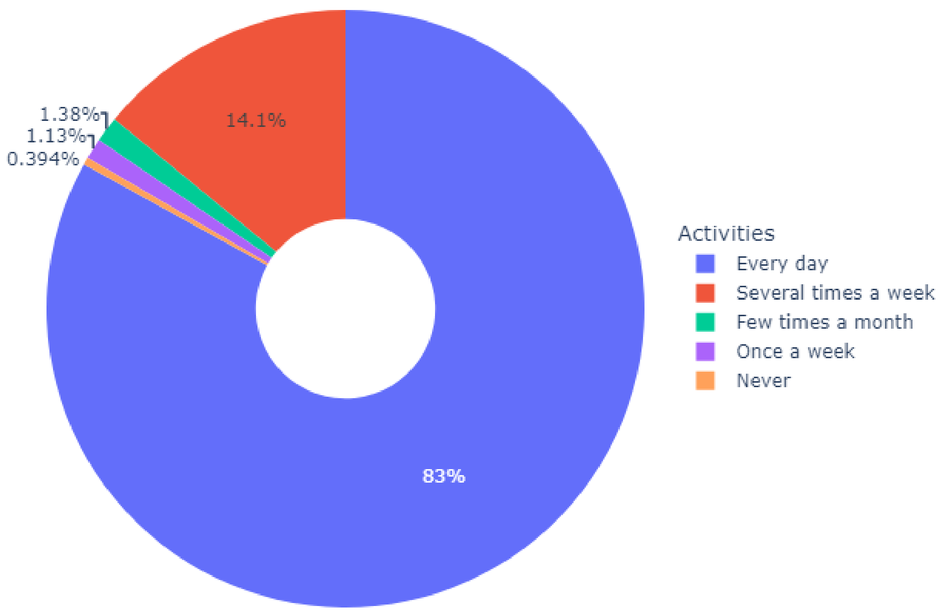

Figure 25: Self-reported monthly activity levels of Mastodon users

The daily activity was divided into nine different options, ranging from a minimum of five minutes to a maximum of more than three hours per day. The majority of participants (approximately 719) reported spending one hour on Mastodon daily, followed by 639 participants who indicated 30 minutes, and 466 who reported 20 minutes a day. The smallest group comprised those who spent just five minutes daily on the platform, with only 22 participants selecting this option.

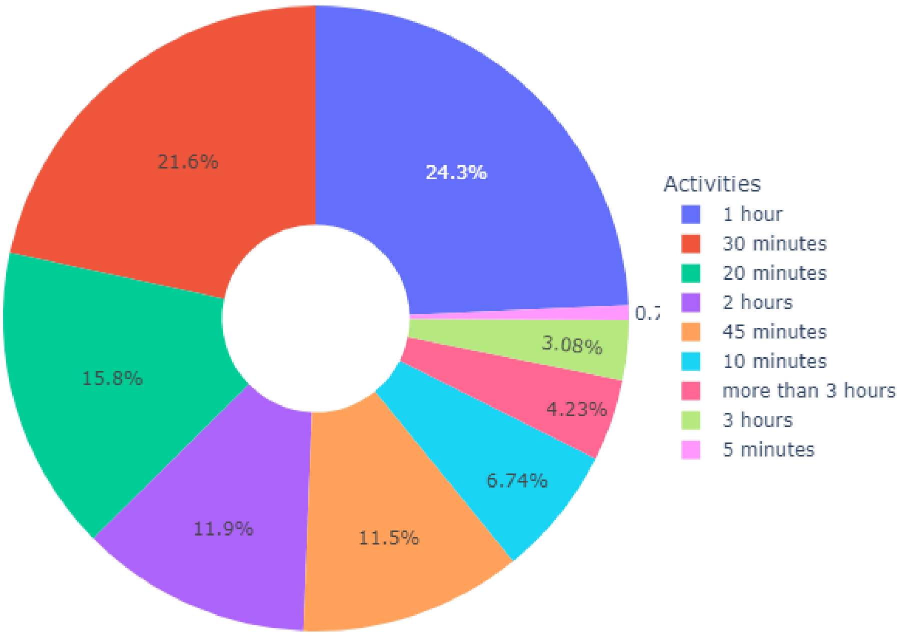

Figure 26: Self-reported daily activity levels of Mastodon users

To prepare the dependent variable for the correlation analysis, both measures of user activity (daily and monthly) were combined into a new column. The 'Every day' column was replaced with the more detailed daily activity data from the second question. The activity levels were then assigned numerical values, with the lowest possible time being given a value of 1 and the highest being given the maximum value. This process resulted in a twelve-point ordinal scale that captures the ranked levels of user activity.

10.5 Independent Variables

This chapter presents the results of the analysis examining the relationship between the independent variables and user activity, the dependent variable of the study. Each subchapter will begin by presenting the distribution of the corresponding factor to establish a clear understanding for the subsequent correlation analysis. Building upon the computational analysis, both Pearson and Spearman correlation coefficients are used to assess linear and monotonic relationships, respectively. In addition, regression analysis is employed to explore the extent to which the independent variables predict user activity levels. The independent variables consist of several theoretically derived factors, which aim to capture not only technical aspects, as shown in the computational results, but also behavioural characteristics. These variables are categorized into three groups: government structures, engagement, and technical factors. Understanding the correlations between these factors and user activity will help provide a foundation for explaining user behaviour on Mastodon.

10.5.1 Technical Infrastructure

The technical infrastructure group includes seven independent factors: information overload, security, uptime, transparency, latency, sustainability, and instance size. With the exception of information overload—which depends on the perception of the user—all of these factors have already been analysed in the computational analysis. However, the survey results in this section will provide better insight into how users perceive the influence of technical infrastructure on user activity.

Instance Size

In this study's survey the majority of participants were on smaller instances, with 659 participants being on instances in the range of 1,001 to 10,000 users. Slightly fewer participants (635) reported being on instances with 100 to 1,000 users. Conversely, only 146 participants indicated that they were on instances with more than 500,000 users. The correlation analysis showed a weak but statistically significant negative correlation between instance size and user activity, with a Pearson correlation coefficient of -0.07 (p<0.05) and a Spearman coefficient of -0.06 (p<0.05). Regression analysis also shows a negative relationship, with a slope of -0.12, suggesting that as instance size increases, user activity decreases slightly. These results are based on a sample of 2,357 responses.

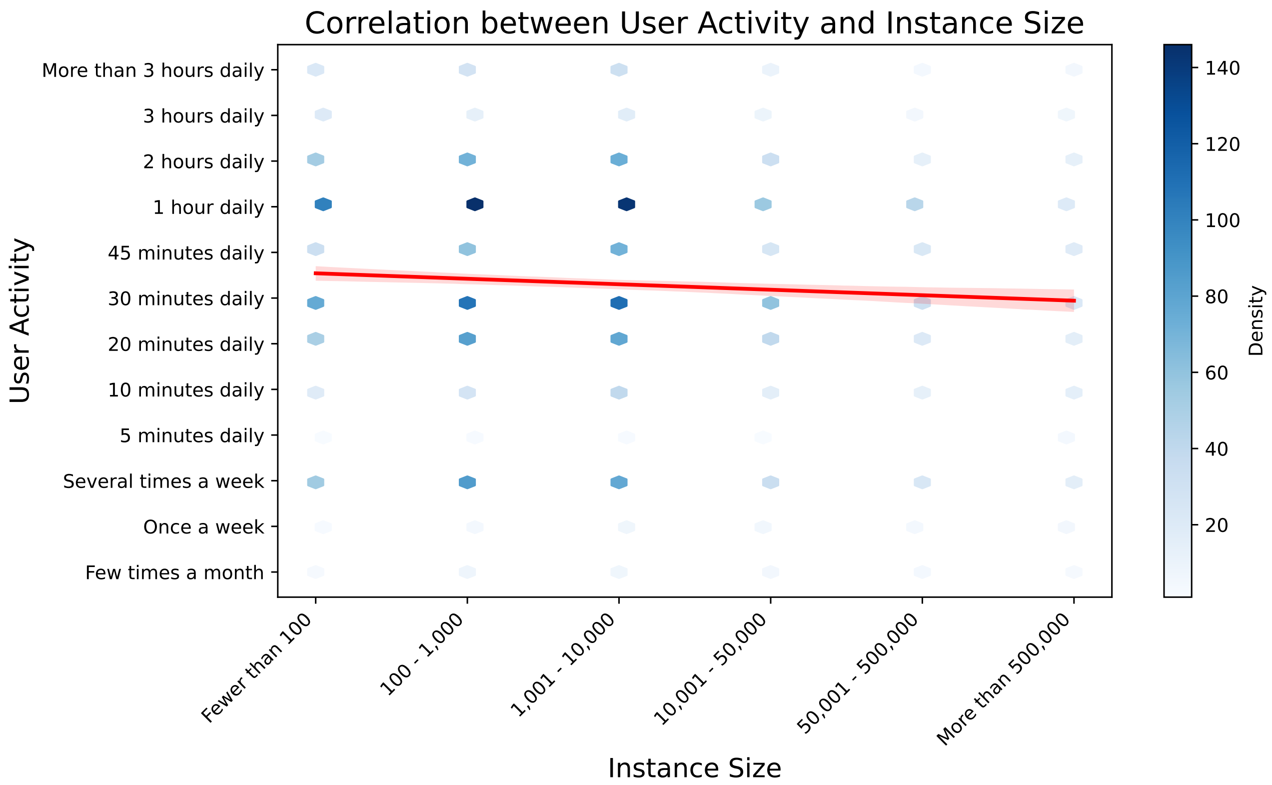

Figure 27: Correlation between instance size and user activity, showing a slight negative trend (Pearson = -0.07, Spearman = -0.06, p < 0.05). The red regression line highlights this trend, indicating that user activity tends to decrease slightly as instance size increases. The darker blue data points in the graph represent higher user density.

Information Overload

The majority of users (65%) were not overwhelmed with the amount of content in their feed and instead have the feeling that the amount of content in their feed is manageable. The Pearson correlation coefficient of 0.001 (p>0.05) and the Spearman correlation coefficient of 0.013 (p>0.05) demonstrate a weak and statistically insignificant relationship between information overload and user activity. The regression analysis indicated similar findings with a coefficient of 0.001 (p>0.05). Therefore, there is no meaningful relationship between information overload and user activity as illustrated in Figure 28.

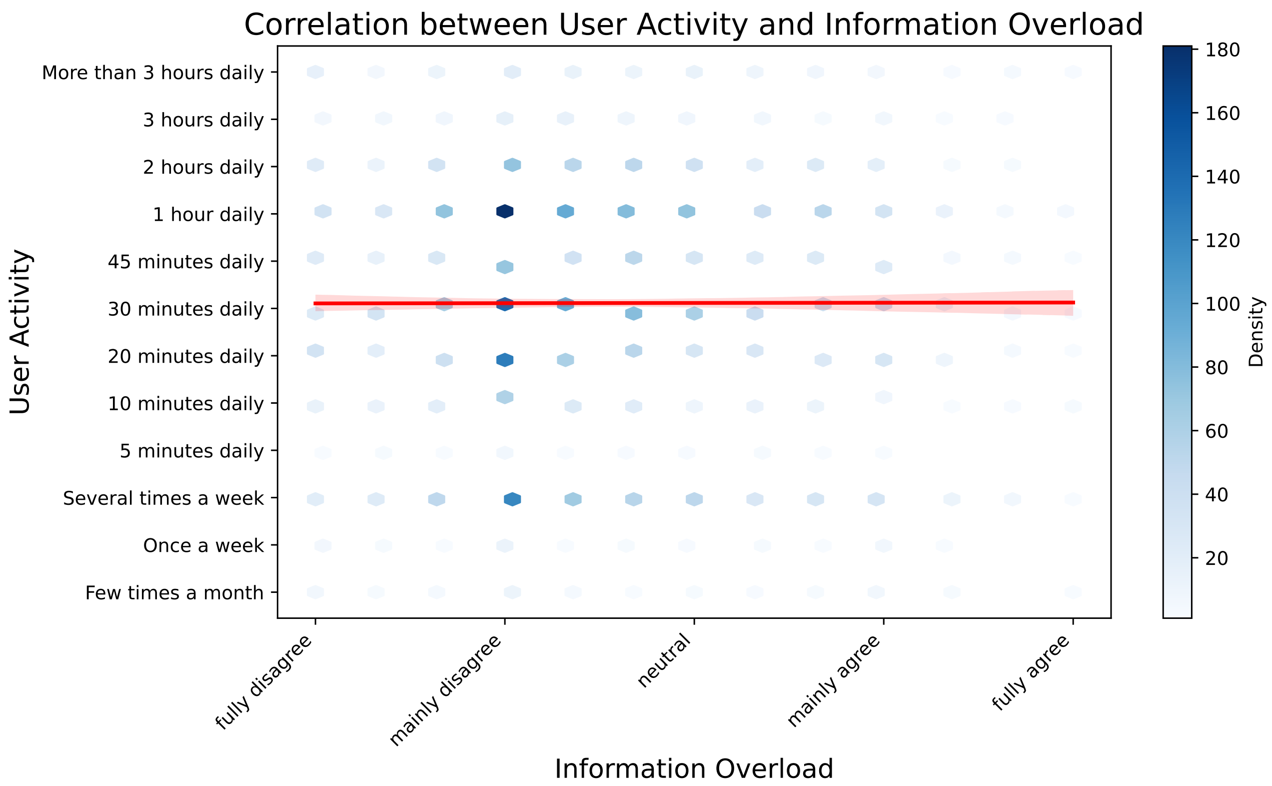

Figure 28: Correlation between information overload and user activity, showing no significant relationship (Pearson = 0.001, Spearman = 0.013, $p > 0.05$). The red regression line also shows that information overload has no meaningful impact on user activity. The darker data points represent higher user density.

Security

75 % of users agree that security implementations on their instance are important to them. However, the correlation analysis shows, that there is only a weak, but statistically significant correlation between security implementations and user activity. The Pearson correlation coefficient of 0.05, (p-value < 0.05) and the Spearman correlation coefficient of 0.05 (p-value< 0.05), indicate a statistically significant yet weak relationship between the independent variable and user activity. The regression analysis further supports this, with a slope of 0.13, indicating a slight increase in user activity as the factor increases. The R-value of 0.05 ($p<0.05$) reaffirm the statistical significance of the rather weak relationship. The results suggest that this factor has a small but significant effect on user activity. The findings are represented in figure 29.

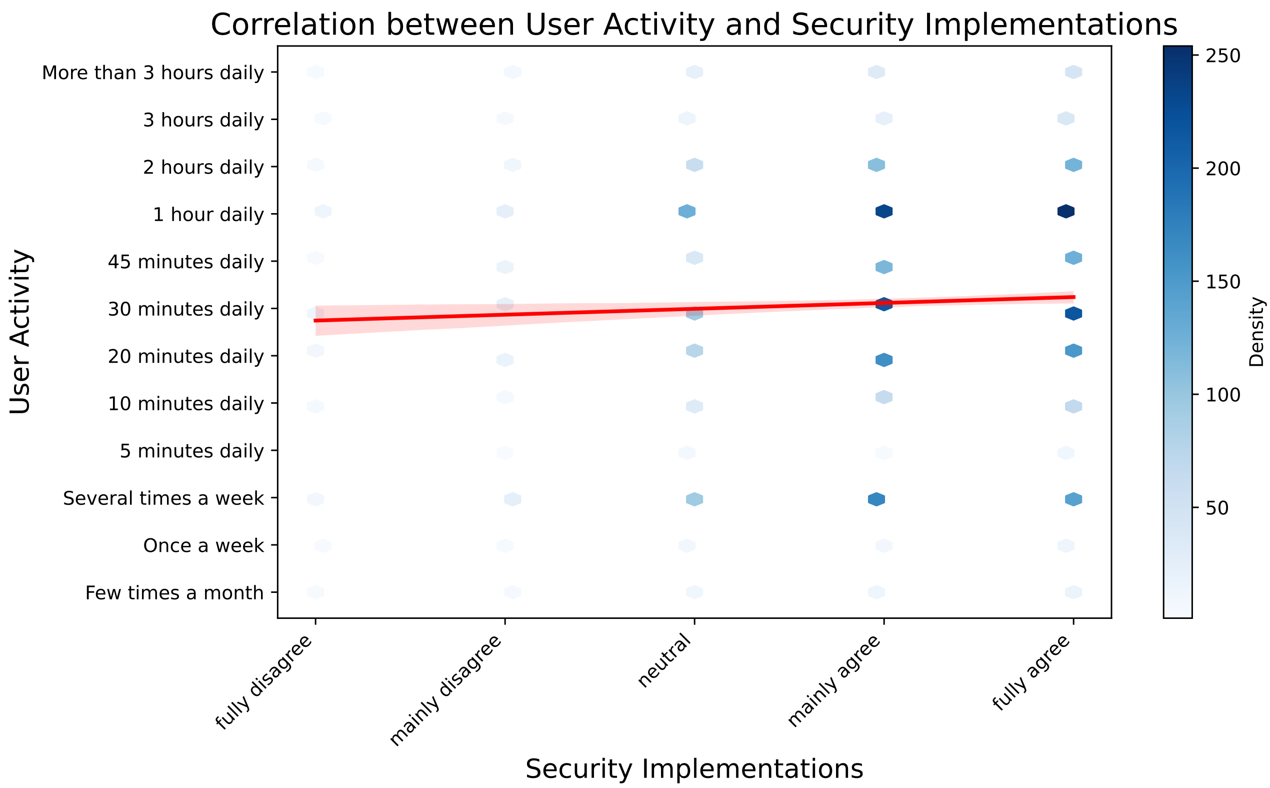

Figure 29: Correlation between security implementations and user activity, showing a slightly positive relationship (Pearson & Spearman = 0.05, p < 0.05). The red regression line in the graph highlights this trend, indicating a slight increase in user activity as more security instalments are implemented on an instance. Darker data points represent higher user density.

Uptime

The majority of the survey participants (97%) agree that their instance is always available. The correlation analysis revealed an inverse relationship between uptime and user activity. The Pearson Correlation revealed a correlation coefficient of -0.04 (p<0.05) and the Spearman a correlation coefficient of -0.06 (p<0.001). The regression analysis points to similar results with an r-value of -0.04. While statistically significant, the relationship is minimal in practical terms, with only a very slight decrease in user activity as uptime increases. A total of 3481 responses were included in this analysis. The results are visualised in Figure 30.

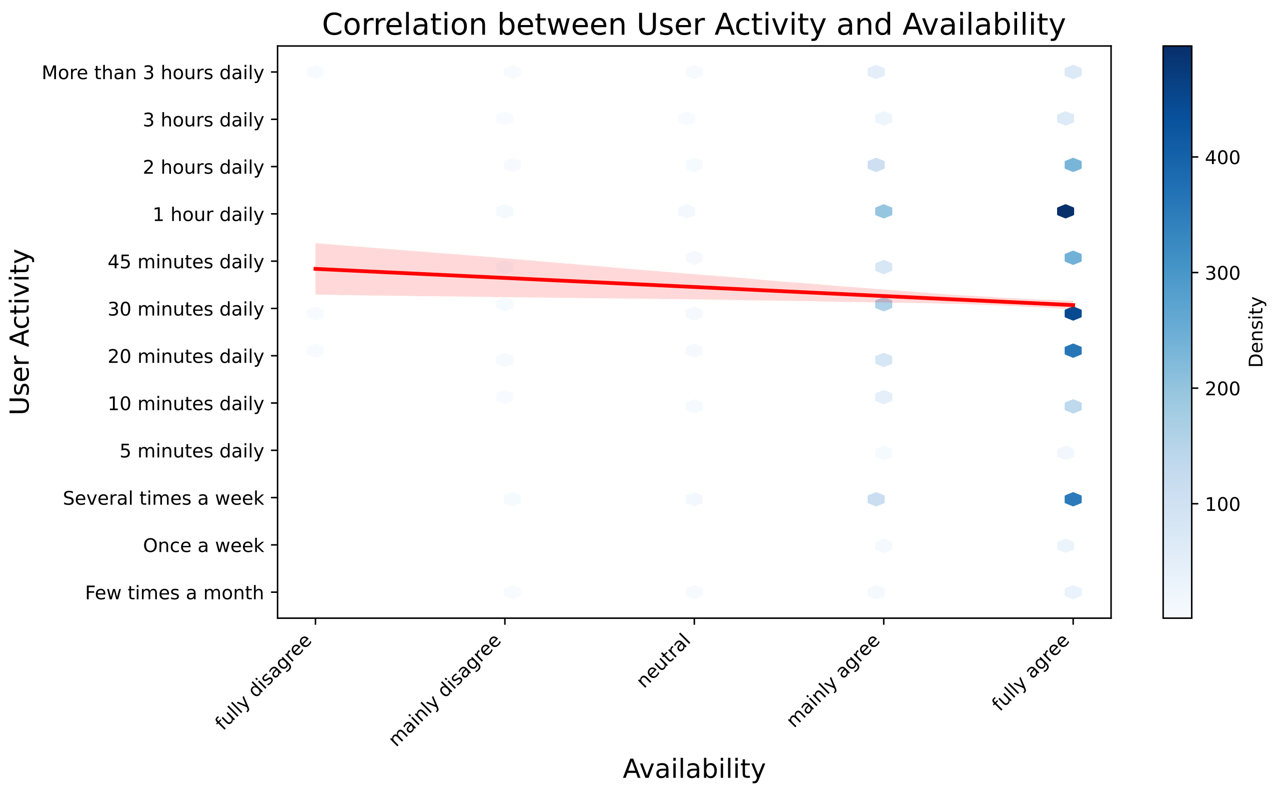

Figure 30: Correlation between instance uptime and user activity, showing a slightly negative relationship (Pearson = -0.04, Spearman = -0.06, $p < 0.05$). The red regression line highlights this trend, indicating a slight decrease in user activity as uptime increases. Darker blue data points represent higher user density.

Transparency

Transparency was primarily assessed through questions about familiarity with and perceived importance of the privacy policies on their Mastodon instance. This variable consisted of two items: familiarity and importance of privacy policies on their instance. The transparency variable yielded a Cronbach's alpha coefficient of 0.74, indicating good internal consistency. Overall, 66% of respondents indicated that they were familiar with the privacy policies of their instance and 72% indicated that these policies were important to them. The correlation analysis revealed a Pearson correlation coefficient of 0.13 ($p<0.001$) and the Spearman correlation coefficient of 0.14 ($p<0.001$). Regression analysis further supported this finding with a coefficient of 0.13 ($p<0.001$). The sample size of this variable were 3,240 responses. The results are visualised in figure 31.

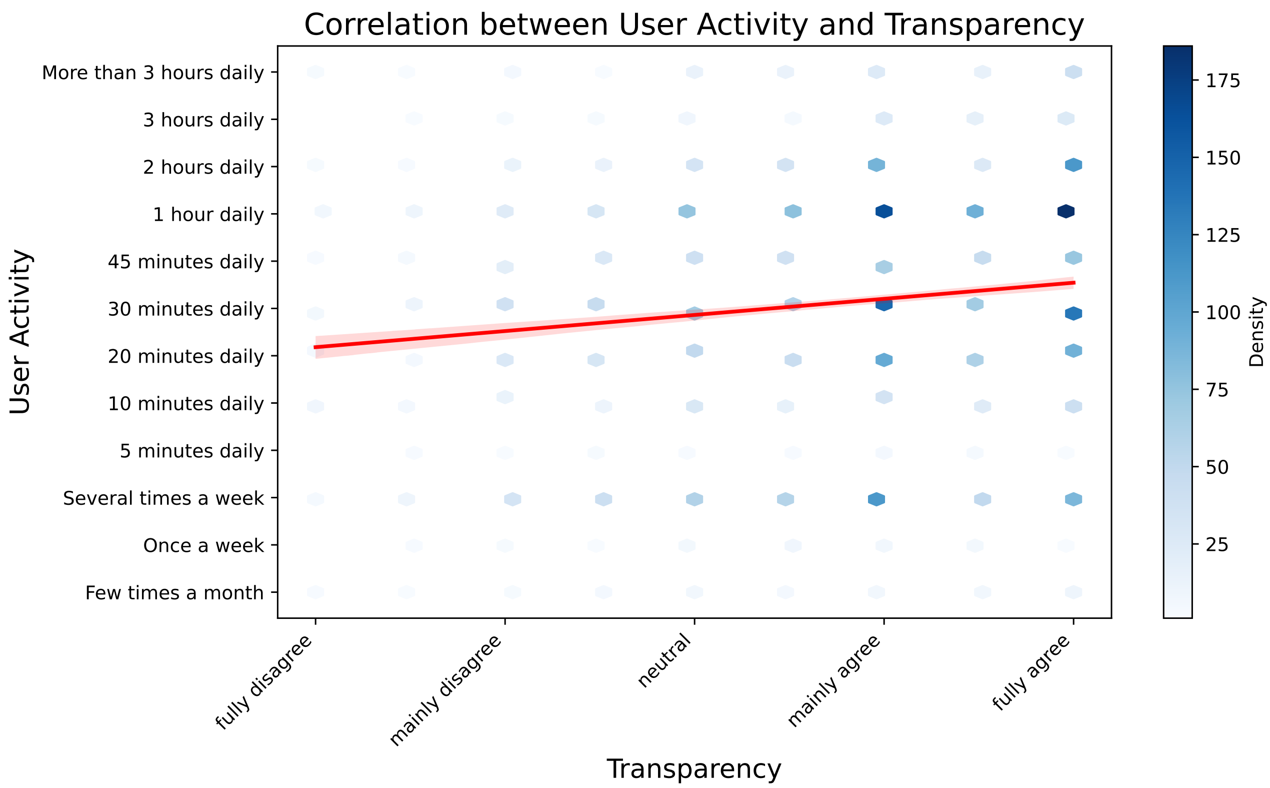

Figure 31: Correlation between transparency and user activity, showing a positive relationship (Pearson = 0.13, Spearman = 0.14, p < 0.001). The red regression line also indicates this trend, with a slight increase in user activity as transparency (in the form of privacy policies) improves. Darker blue data points represent higher user density, with a high proportion of users agreeing that privacy policies are important to them.

Latency

When Mastodon users were asked if they experienced latency on the platform, only 9% agreed. This suggests that the majority of Mastodon users do not experience latency problems, which is consistent with the results of the computational analysis. Correlation analysis revealed a Pearson correlation coefficient of -0.06 (p<0.001), a Spearman rank correlation coefficient of -0.07 (p<0.001) and the regression analysis revealed a negative coefficient of -0.06 (p<0.001). These results suggest that as latency decreases (i.e. users experience fewer delays), there is a slight decrease in user activity. The sample size was 3,473 participants. The results are shown in Figure 32.

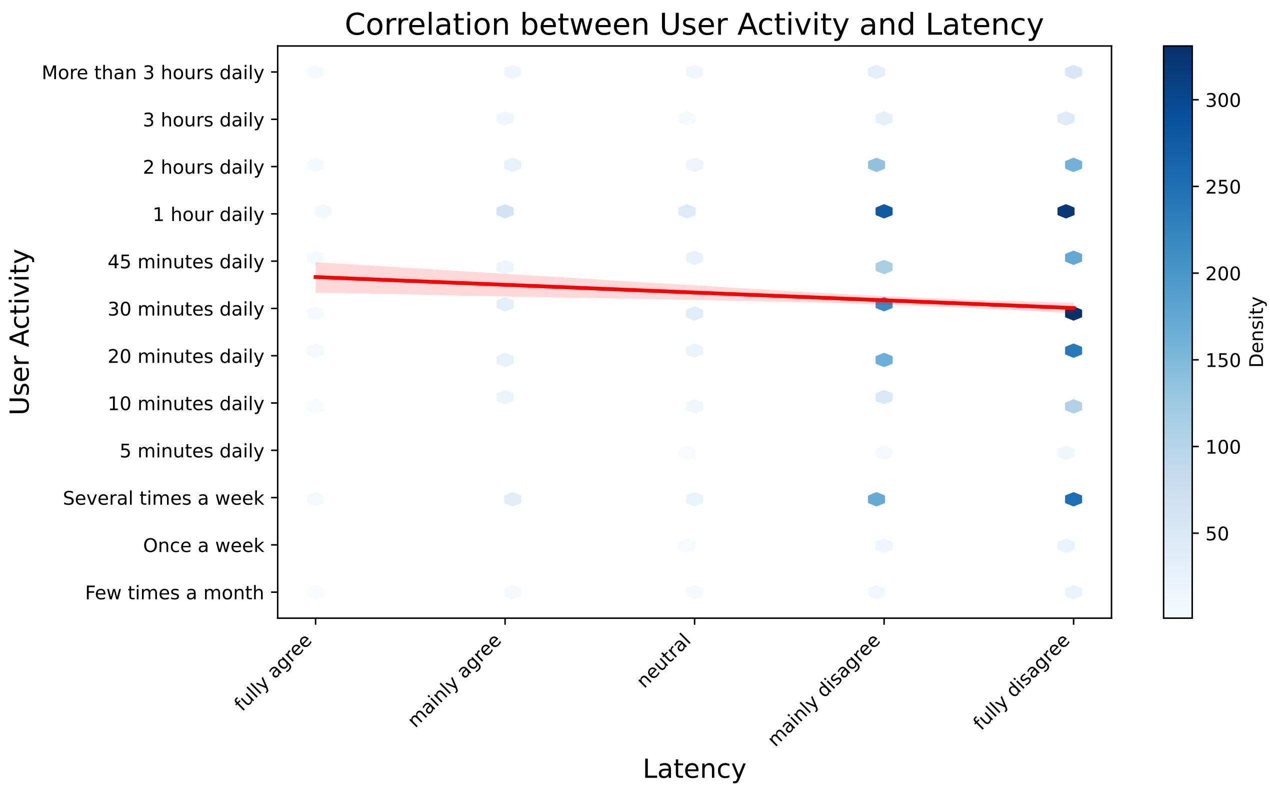

Figure 32: Correlation between latency and user activity, showing a weak, negative relationship (Pearson = -0.06, Spearman = -0.07, $p < 0.001$). The red regression line further supports this trend, indicating a slight decrease in user activity as latency decreases. Darker blue data points represent higher user density, showing that most of the participants experience no latency.

Sustainability

The majority of survey participants (55%) indicated that sustainability on their instance, specifically in the form of being powered by green, sustainable energy, is important to them. However, the Pearson and Spearman correlation analyses and the Regression analysis revealed weak and non-statistically significant results, suggesting that the perceived importance of sustainability does not meaningfully correlate with user activity. The analysis incorporated a total of 3384 responses. Figure 33 illustrates the relationship between sustainability and user activity.

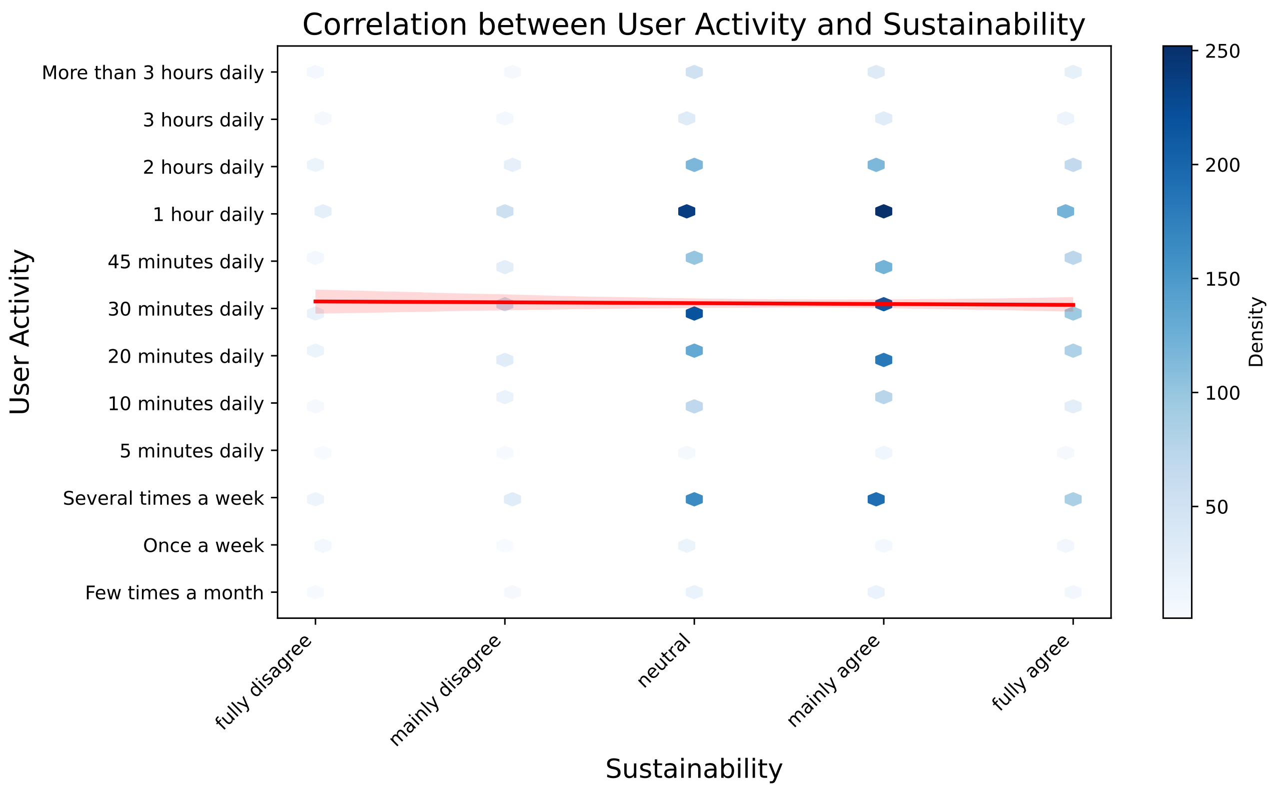

Figure 33: Correlation between sustainability and user activity, showing no statistically significant relationship. The red regression line further supports that the perceived importance of sustainability does not meaningfully correlate with user activity. Darker blue data points represent higher user density.

10.5.2 Governance Structures

The construct of governance structures encompasses two key components: community guidelines and moderation policies. These factors collectively define the rules, norms, and enforcement mechanisms that shape user behaviour and interactions within Mastodon. This chapter explores the results related to both community guidelines and moderation policies, providing insights into their correlation with user activity.

Community Guidelines

To assess the Community Guidelines factor, it was essential to first ensure that all participants were familiar with the community guidelines on their instance. Therefore, a filter question was included, asking whether participants were familiar with the community guidelines on their instance. Of the 3,347 participants, 2,681 indicated familiarity with their community guidelines. The correlation between community guidelines familiarity and user activity was also analysed, treating responses as binary ('No' and 'Yes'). This analysis resulted in positive Pearson and Spearman correlation coefficients of 0.14, both statistically significant.

However, this binary measurement was not used in the main results, as a more robust 5-item scale was considered to be more reliable. The community guidelines factor was then constructed from four items, with a Cronbach's alpha of 0.83, indicating good internal consistency. These items were combined into an indexed and averaged score. Of the 2,529 participants who were familiar with their community guidelines and answered all 4 items in this set, 794 (31%) strongly agreed that their instance guidelines were clear, promoted inclusivity, were easy to understand and access, and addressed the most important issues to them. The correlation analysis revealed a statistically significant, positive relationship between the community guidelines factor and user activity. The Pearson correlation coefficient was 0.11, while the Spearman correlation coefficient was 0.13. Regression analysis produced similar results, with a coefficient of 0.11. In all cases, p-values were below 0.001, confirming the statistical significance of the findings. The results are visualized in Figure 34.

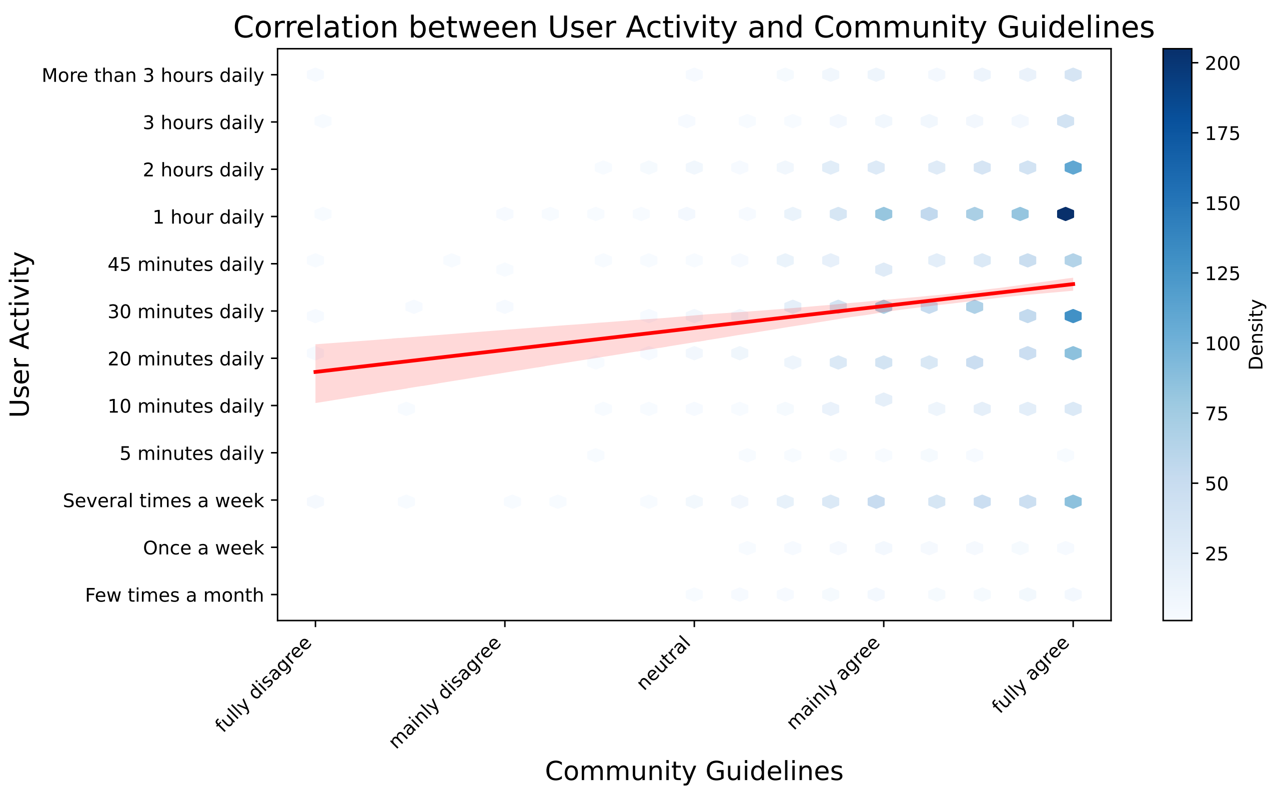

Figure 34: Correlation between community guidelines and user activity, showing a positive relationship (Pearson = 0.11, Spearman = 0.13, $p < 0.001$). The red regression line supports that clearer, more inclusive, and accessible community guidelines are associated with slightly higher user activity levels. Darker blue data points represent higher user density, which indicates that most of the survey participants agree that community guidelines are important to them.

Moderation Practices

The moderation practices variable originally consisted of four items; however, due to low internal consistency (as discussed in the data processing chapter), two items were removed. The final moderation practices variable included two items that measured whether participants felt moderators were taking action to enforce community guidelines and whether they responded quickly to reports. Of the respondents, 41% fully agreed that moderators were both enforcing community guidelines and replying promptly to reports. Correlation analysis showed a positive, statistically significant relationship between the moderation practices variable and user activity. The Pearson correlation coefficient was 0.11, while the Spearman coefficient was 0.13, both with p-values below 0.001. These findings were confirmed by the regression analysis, which produced a statistically significant coefficient of 0.11 ($p<0.001$). Figure 35 shows a Hexbin plot illustrating the relationship between moderation practices and user activity, with a colour gradient representing the density of responses. A red regression line overlays the plot, highlighting a positive trend, indicating that stronger moderation practices are associated with slightly higher levels of user activity.

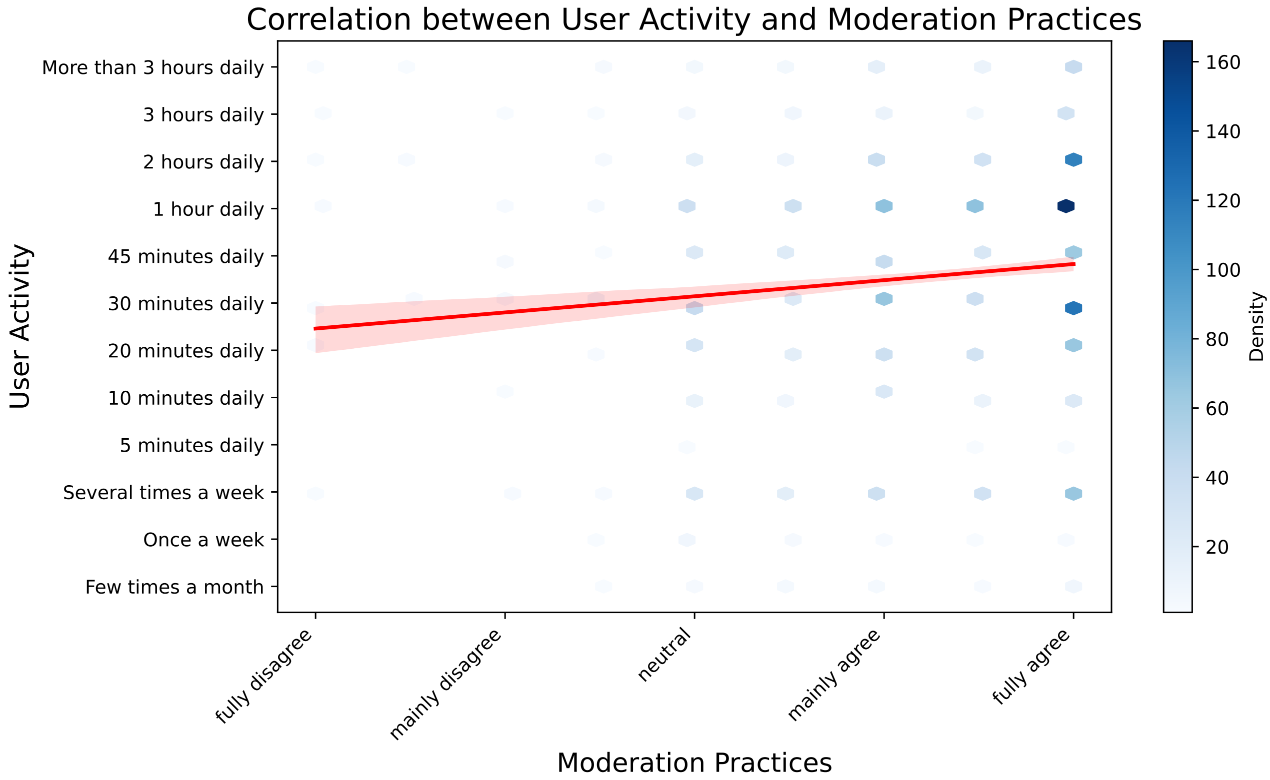

Figure 35: Correlation between moderation practices and user activity, showing a statistically significant positive relationship (Pearson = 0.11, Spearman = 0.13, $p < 0.001$). The red regression line suggests that instances with actively enforced community guidelines tend to have slightly higher user activity levels. Darker areas indicate a higher concentration of data points showing that most of the users agree that moderation practices are actively enforced on their instance.

10.5.3 Engagement

Engagement can be measured in several ways. In this thesis we differentiate between active engagement, where a user is online, replying, boosting and favouring other toots. Whereas passive engagement refers to only being online without any active engagement, which can be compared to watching television [64]. The engagement hypotheses suggest that higher engagement leads to higher user activity levels. In this thesis, engagement is categorized into three groups: (1) Active contribution, where a user proactively publishes their own toots; (2) inbound engagement, referring to the interactions a user receives on their published toots; and (3) outbound engagement, where a user interacts with the toots of other users on their feed, such as by replying to, boosting, or favouring toots. This chapter presents the correlation between various types of engagement and user activity levels.

Active Contribution

The majority of Mastodon users reported that they toot sometimes (35%) or often (24%), while smaller proportions reported that they toot very often (14%) or never (5%). Correlation analysis showed a moderate positive correlation between user activity and active contribution on Mastodon. The Pearson and Spearman correlation analysis and the regression analysis revealed positive, statistically significant coefficients of 0.38 ($p < 0.05$). This means that the more often a user toots, the more active they are on Mastodon. The large sample size of 3,144 gives confidence in the robustness of this result. The data points and regression line are shown in Figure 36.

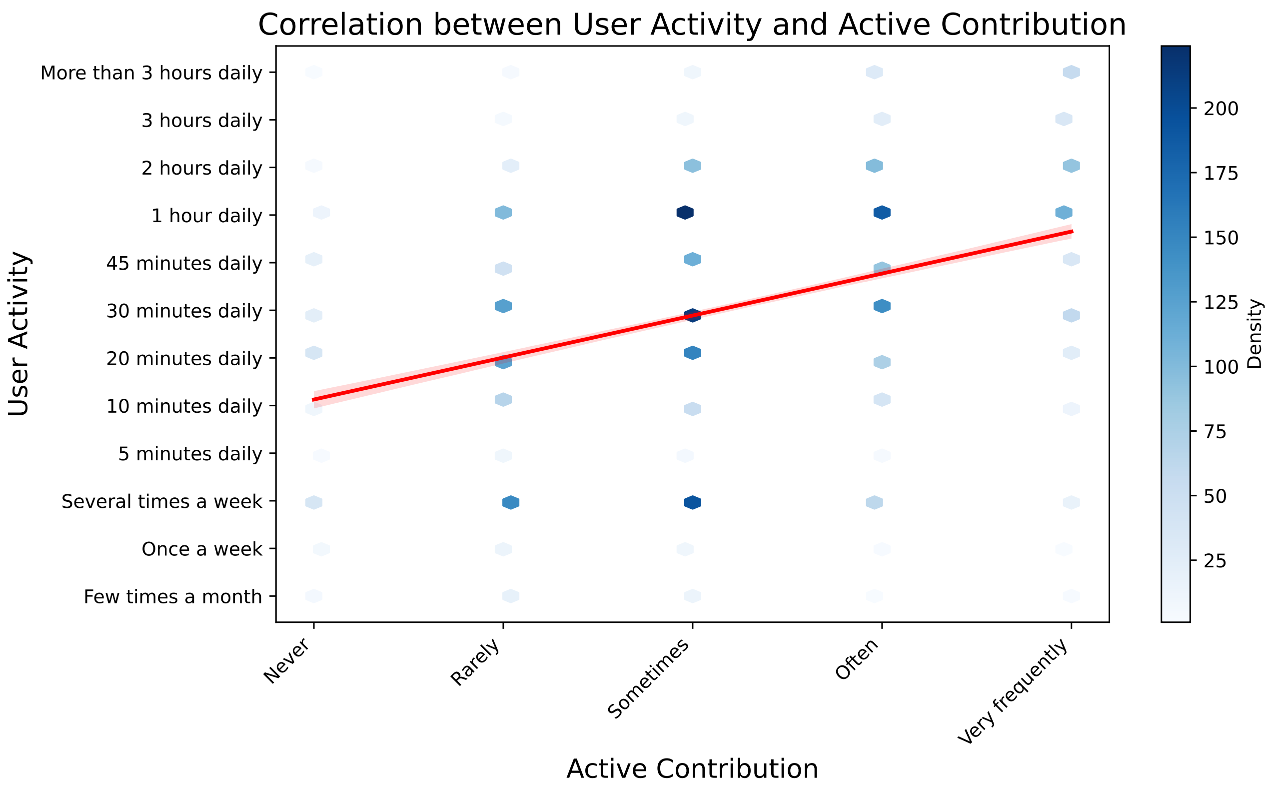

Figure 36: Correlation between active contribution frequency and user activity, showing a moderate, positive relationship (Pearson = 0.38, Spearman = 0.38, $p < 0.05$). The red regression line highlights this trend, indicating that users who toot more frequently tend to have higher overall activity levels on Mastodon. Darker data points represent areas of higher response density, showing that most users toot sometimes or often.

Inbound Engagement

Inbound engagement in this thesis refers to the engagement a user receives on their posts. This variable was consisted of three items: (1) satisfaction with the amount of engagement, (2) wanting to receive more engagement, and (3) actually receiving engagement when tooting. The internal consistency was measured with a Cronbach's alpha of 0.74. Most respondents (63%) indicated that they typically receive some engagement when they toot, while 37% expressed that they would like to receive more engagement on their toots. Correlation analysis showed a weak, positive relationship between engagement on Toots and user activity with a statistically significant Pearson correlation coefficient of 0.20, a Spearman correlation coefficient of 0.19 and regression analysis coefficient of 0.20 ($p<0.001$). The sample size was 3.217 and the results are visualised in Figure 37.

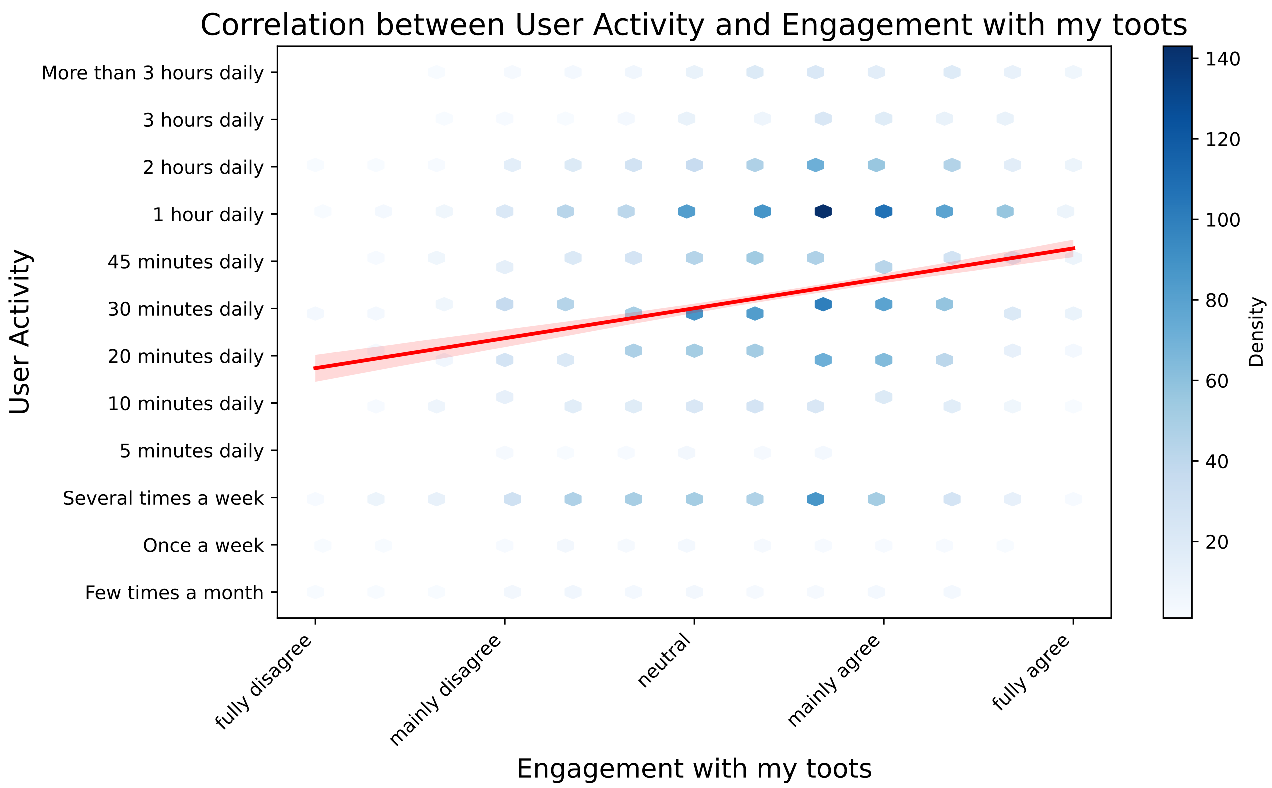

Figure 37: Correlation between inbound engagement and user activity, showing a weak but statistically significant positive relationship (Pearson = 0.20, Spearman = 0.19, p < 0.001). The red regression line supports this trend, suggesting that users who receive more engagement on their toots tend to have slightly higher activity levels. Darker data points represent areas with more user responses.

Outbound Engagement

Outbound engagement is the interaction of a user with the content/toots of other users. This variable consists of three items: (1) how often a user replies to (2) favourites and (3) boosts other toots. The internal consistency was measured with a Cronbach's alpha of 0.74. Comparing replying to toots, favouring and boosting other toots, users most often favour toots (36%), then boost toots (22%) and only 8% of users very frequently reply to other toots. The analysis revealed a moderate positive relationship between outbound engagement and user activity. The Pearson correlation analysis and the regression analysis revealed a coefficient of 0.38 (p<0.001), and the Spearman correlation one of 0.40 (p<0.001). These findings are based on a sample size of 3,529 respondents and are presented in Figure 38.

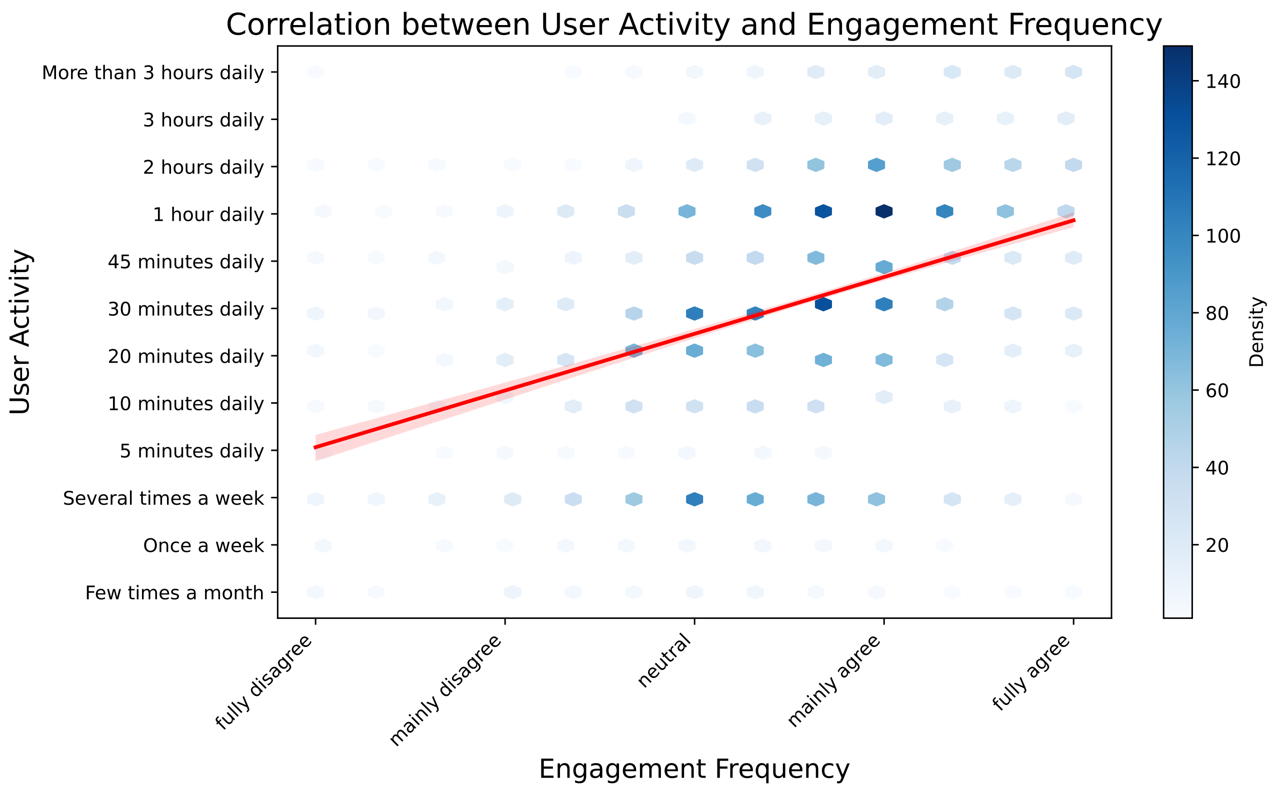

Figure 38: Correlation between outbound engagement and user activity, showing a moderate positive relationship (Pearson = 0.38, Spearman = 0.40, both $p < 0.001$). The red regression line indicates that users who interact more frequently with others' toots—by replying, favouring, or boosting—tend to have higher activity levels. Darker data points represent areas of higher response density, indicating that most users engage at moderate to high levels on average.

10.6 Additional factors

This chapter contains additional factors that were not previously derived from theory, but emerged from the analysis of the collected survey data.

10.6.1 Years on Mastodon

Nearly half (47%) of the respondents joined Mastodon in 2022, a period marked by a significant surge in the platform's popularity. User activity shows a positive correlation with the number of years spent on Mastodon. This suggests that users who joined Mastodon earlier tend to spend more time on the platform. The Pearson correlation analysis and the regression analysis indicate a weak, positive relationship, with coefficients of 0.16 ($p<0.001$), while the Spearman coefficient is 0.15 ($p<0.001$). The results are based on 3,454 responses. The findings of the analysis are presented in Figure 39.

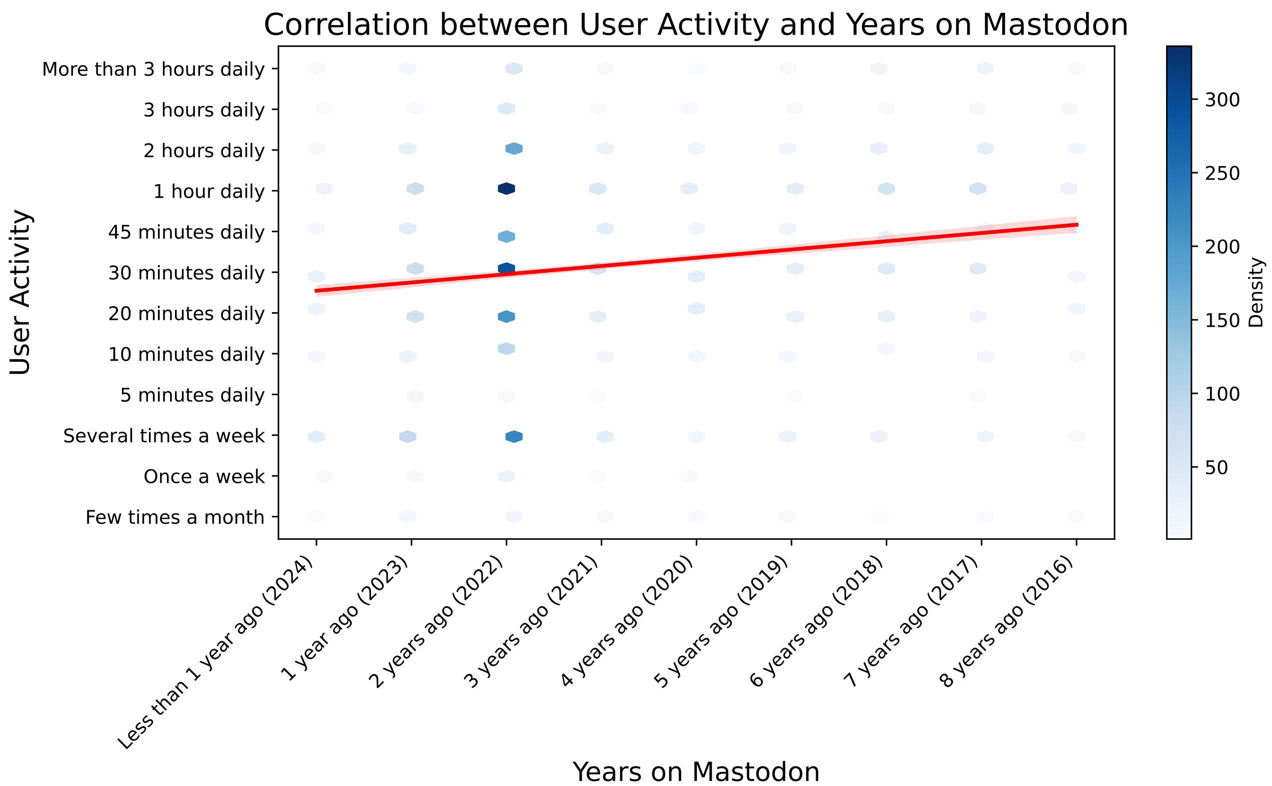

Figure 39: Correlation between years on Mastodon and user activity, showing a weak but statistically significant positive relationship (Pearson = 0.16, Spearman = 0.15, $p < 0.001$). The red regression line suggests that users who have been on Mastodon for a longer period tend to be more active. Darker data points represent higher response density, showing that most users joined 2 years ago.

10.6.2 Exploring new factors

At the end of the survey, participants had the opportunity to respond to an open-ended question regarding which factors they believed influenced their user activity on Mastodon. A total of 2,441 participants provided responses, which led to an exploratory analysis as no predefined variables or hypotheses were established beforehand. The dataset was analysed to identify patterns, correlations, or clusters without relying on predetermined assumptions. As a first step, a frequency analysis of the most mentioned words was conducted to uncover recurring themes and trends in the responses. For this purpose, libraries such as sklearn and wordcloud were employed. The sklearn library, specifically the CountVectorizer tool, was used to process the text data by filtering stopwords, vectorizing the text, and calculating word frequencies. Finally, the wordcloud library was used to visually represent the most frequently mentioned words in the dataset, in order to provide an overview of key terms that emerged from the responses. The visualisation is presented in Figure 40.

Figure 40: Additional factors influencing user activity on Mastodon, as indicated by survey participants (visualized as a word cloud)

The most frequently mentioned words were people (346), time (192), content (163), community (159), interesting (159), mastodon (155), "Zeit" (153), social (150) and like (148). While this may assist with the analysis, it does not provide a comprehensive overview. The subsequent stage of the analysis entailed the investigation of the frequency with which multiple words co-occurred, thus leading to the identification of salient topics including social media, interesting people, free time, people follow, interesting content, "verfügbare Zeit", similar interests, available time and like minded.

After the frequency analysis, the next step involved applying topic modelling using the BERTopic framework to gain deeper insights into the open-ended survey responses. This process began by cleaning the dataset, removing missing or irrelevant entries to ensure accurate analysis. A multilingual BERTopic model was then employed to extract meaningful topics from the text data. The model was fitted to the responses, and topics were assigned to each entry in the dataset. Detailed information about each topic, including the most representative words and their weights, was generated and saved in a CSV file for further reference. Additionally, the frequency distribution of topics across the dataset was calculated and saved separately. To enhance understanding, the relationships between topics were visualized using BERTopic's built-in visualisation tools. The topic names were then assigned manually for greater interpretability. This comprehensive topic modelling approach provided a structured way to analyse the qualitative data, complementing the initial word cloud analysis by uncovering recurring themes and their significance.

The analysis revealed several recurring topics influencing user activity on Mastodon. These included the social media alternatives, the presence of an active community, and interest in news and current events. Other significant factors were availability of time, a friendly and respectful environment, and the need to combat boredom. Users also valued the platform's ad-free experience, mood-related factors, and the ability to connect over common interests, curate their own feeds, the absence of algorithms and the importance of shared values. Finally, ChatGPT, a large-scale AI language model developed by OpenAI, was employed to analyse batches of 300–500 rows of text data. [4]. The analysis aimed to identify the 10 most common topics mentioned by respondents. Key themes included community and connections, availability of time, interesting content, politics and social issues, algorithm-free and ad-free experiences, positive, friendly, and respectful interactions, diversity and inclusivity, creative content sharing, user engagement, and the benefits of decentralization and open platforms, among others.

11 Discussion

This chapter discusses the findings of this study, which explored user activity on Mastodon and the factors that correlate with it. The results are interpreted in relation to the research questions and existing literature to provide deeper insights into the dynamics of user behaviour on decentralised social media platforms. The factors that were hypothesised to correlate with user activity are engagement, governance structures and technical infrastructure. Three approaches were initially used for correlation analysis - Pearson's correlation, Spearman's rank correlation and regression analysis. The results presented here are based on Spearman's correlation, as it is most appropriate for the ordinal nature of the survey data [75]. Spearman's method is also more robust to outliers, which is crucial given the characteristics of the API data, where many data points were observed but outliers could potentially distort the results. The results of Pearson's correlation and regression analysis are included in the results section above.

11.1 Overview of Findings

This section presents a structured summary of the key findings, focusing on the discussion of the hypotheses. It includes a table and a graph summarizing the results, which will be further examined in the discussion. The results can be summarised as follows:

- **Governance Structures are important:** Transparent, inclusive community guidelines and responsive moderation show a positive correlation with user activity. This implies that instances fostering an inclusive, well-regulated, and clearly structured environment will also have more active users.
- **Technical infrastructure does not strongly correlate with user activity:** Uptime, latency, sustainability, and information overload yield conflicting or non-significant results. These findings suggest that while a stable and well-maintained infrastructure is crucial for the functionality of an instance, it has little direct impact on user activity levels.
- **Transparency matters:** Transparency shows a positive correlation in both the survey and the computational analysis, indicating that privacy policies are important to Mastodon users. This suggests that instances with higher transparency exhibit higher activity levels.

[4] https://chatgpt.com/

- **Engagement is the strongest predictor of user activity:** Users who post frequently, interact with others, and receive engagement on their posts tend to be more active than passive users who primarily scroll through their feeds. This suggests that admins and moderators should actively encourage engagement within their instance, as higher interaction levels are associated with increased user activity.

- **Smaller instances foster higher engagement:** Smaller communities demonstrate higher activity rates in both the user survey and computational analysis, supporting prior findings that instance size influences activity levels. This suggests that smaller instances may cultivate a stronger sense of community and interaction, fostering more active participation. In contrast, larger instances may lead to decreased user activity, contributing to user discontinuation of Mastodon use.

Figure 41 provides an overview of the survey results based on Spearman correlation analysis. It highlights the strong correlation between engagement and user activity while also presenting other key factors influencing user activity on Mastodon. Table 10 presents a comparison between the proposed hypotheses and the corresponding findings, highlighting the extent to which each hypothesis is supported by the study's results.

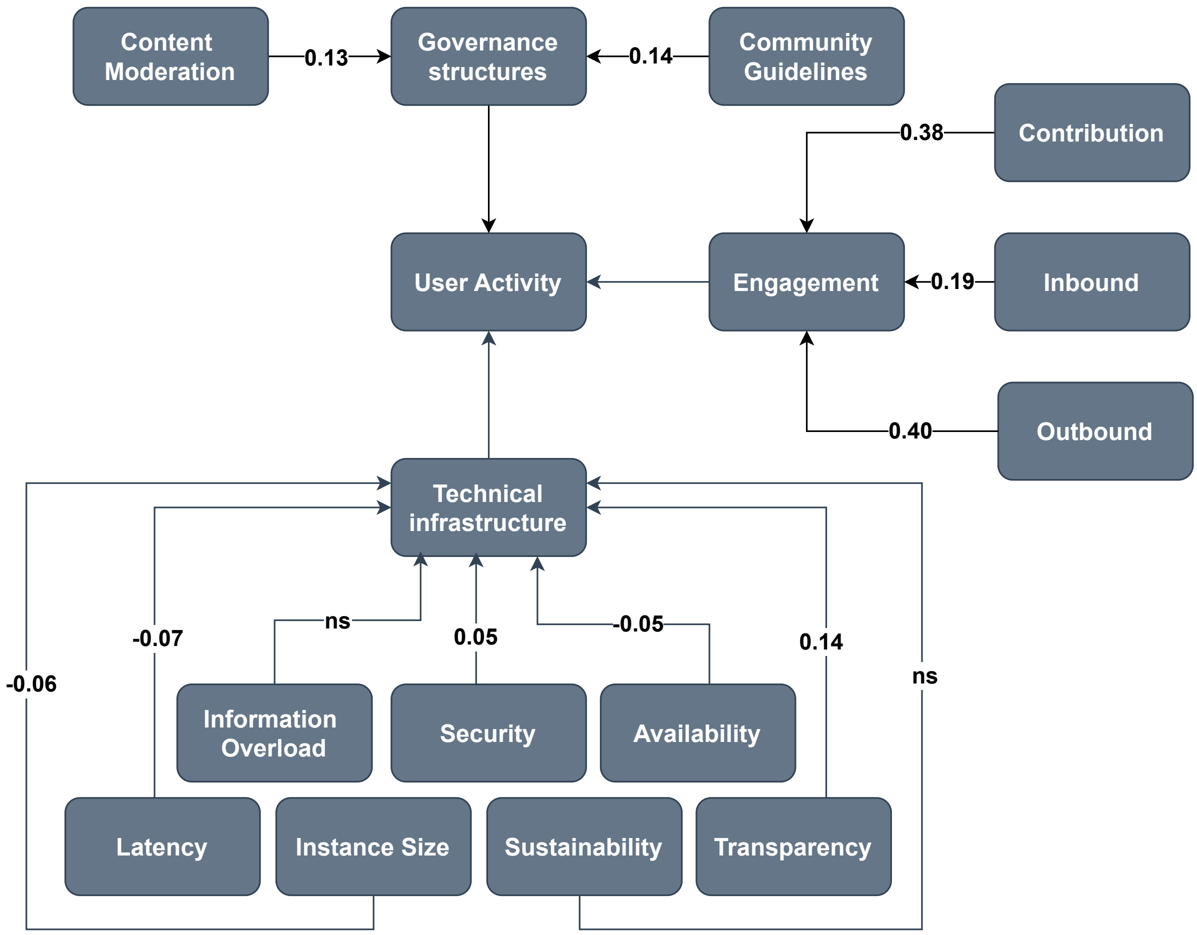

Figure 41: Overview of user survey results and the correlation between technical factors, engagement, governance structures, and user activity. The figure illustrates the relationships between these key variables based on Spearman correlation coefficients, with non-significant correlations (p > 0.05) indicated as ns in the visualisation.

Hypothesis	Finding
H1a: Instances with well-defined, transparent and inclusive community guidelines foster higher user activity rates.	**Supported** (Survey: Spearman = 0.14 ($p < 0.001$))
H1b: Instances with responsive and proactive content moderation demonstrate higher user activity rates.	**Supported** (Survey: Spearman = 0.13 ($p < 0.001$))
H2a: Instances with higher uptime exhibit higher user activity levels.	**Conflicting Results** (Survey: Spearman = -0.06 ($p < 0.001$), Computational: Spearman = 0.25 ($p < 0.001$))
H2b: Instances with lower latency demonstrate higher user activity levels.	**Conflicting Results** (Survey: Spearman = -0.07 ($p < 0.001$), Computational: Non-Significant)
H2c: Instances with enhanced security measures exhibit higher user activity levels.	**Conflicting Results** (Survey: Spearman = 0.05 ($p < 0.05$), Computational: Spearman = -0.07 ($p < 0.05$))
H2d: Instances with greater transparency, including clear privacy policies and terms of service, foster higher user activity levels.	**Supported** (Survey: Spearman = 0.14 ($p < 0.001$), Computational: Spearman = 0.17 ($p < 0.001$))
H2e: Information overload leads to lower user activity levels.	**Not supported** (No significant correlation)
H2f: Environmentally sustainable instances have higher user activity levels.	**Not supported** (No significant correlation)
H2g: Smaller instances exhibit higher user activity levels.	**Supported** (Survey: Spearman = -0.06 ($p < 0.05$), Computational: Spearman = -0.64 ($p < 0.001$)
H3a: Users who proactively toot demonstrate higher levels of user activity.	**Supported** (Survey: Spearman = 0.38 ($p < 0.001$))
H3b: Users who engage with the content of other users exhibit higher levels of user activity.	**Supported** (Survey: Spearman = 0.40 ($p < 0.001$))
H3c: Users who receive higher engagement on their toots exhibit higher levels of user activity.	**Supported** (Survey: Spearman = 0.19 ($p < 0.001$))

Table 10: Summary of hypothesis testing and their findings

11.2 Detailed discussion of findings

This chapter discusses the three research questions and their associated hypothesis and compares them to the findings of this study. The first research question was formulated as follows:

- **RQ1:** How do governance structures impact user activity on Mastodon instances?

Two hypotheses were derived from this research question. The first examined community guidelines and was formulated as follows:

- **H1a:** Instances with well-defined, transparent and inclusive community guidelines foster higher user activity rates.

Xu et al. [92] suggested that community policies have a limited effect on users' intention to switch platforms, based on their study of Facebook users. They argue that over time, social media platforms have adopted similar policies, which may act more as a hygiene factor rather than a motivating one. However, the significant shift of many Twitter users to Mastodon following changes in ownership and content moderation policies

[85] suggests that Mastodon users may be more sensitive to community guidelines than users of commercial social media platforms. The findings from this study support this idea, showing a positive relationship between clear, transparent, and inclusive community guidelines and user activity. This is indicated by a Spearman correlation coefficient of 0.14 and a p-value below 0.001. The positive correlation between community guidelines and user activity suggests that when users understand, identify with and feel included on their instance this would motivate them to interact more frequently with their instances. Furthermore, it is worth considering that these results reflect Mastodon's alternative approach to governance structures, where users have the opportunity to choose their instance along with the instance policies, or to create their own instance based on their values. The sense of belonging to a community requires, among other factors, the development of inclusive policies and the creation of spaces that actively promote and address the different dimensions of belonging [10]. This highlights the important role of inclusive policies in both offline and online communities, shaping community dynamics and user behaviour.

The second hypothesis, formulated in response to the first research question, was:

- **H1b:** Instances with responsive and proactive content moderation demonstrate higher user activity rates.

Prior research has demonstrated that moderation measures can significantly reduce posting activity [42]. Conversely, the presence of engaging and experienced moderators who effectively enforce community policies and rules has been shown to enhance the perceived sociability of an online community [63]. These findings highlight the dual impact of moderation practices on user behaviour and suggest that moderators may play a significant role in shaping user activity on Mastodon. The user survey conducted for this thesis on Mastodon supports this assumption, demonstrating a Spearman's correlation coefficient of 0.13 with a highly significant p-value of < 0.001. The moderation practices examined in this study focused on whether users felt that moderators actively enforced community guidelines and whether they responded promptly to user reports. For Mastodon, with its decentralised structure and user-governed instances, moderators fulfil a distinctive role as both enforcers of policies and facilitators of a positive user experience. The correlation, however, is only modestly strong, suggesting that while moderation is an important factor, it is unlikely to be the sole determinant of user activity. Therefore, the second research question was suggested as follows:

- **RQ2:** How does the technical infrastructure affect user activity on Mastodon instances?

The findings of Xu et al. (2014) [92] demonstrated that technical quality does not significantly influence user retention. However, it was assumed that the decentralised nature of Mastodon might lead to a different outcome. The technical infrastructure term of the second research question encompasses various items, including information overload, security, uptime, latency, sustainability, transparency, and instance size. It was hypothesized that most of these factors positively correlate with user activity, with the exception of instance size, latency and information overload, which may have a negative effect on user activity. To evaluate these hypotheses, technical infrastructure variables were assessed through both computational analysis and a user perception survey, offering a comprehensive understanding of their impact. The computational analysis distinguished between large and small instances; however, for the purpose of this discussion, only large instances (those with over 100 users) will be considered to ensure comparability of results. The results of the computational analysis are summarised in Figure 42.

- **H2a:** Instances with higher uptime exhibit higher user activity levels.

Uptime is a critical component of the technical infrastructure on social networking sites, impacting platform reliability and user experience. This hypothesis posits that instances with higher uptime are likely to demonstrate increased user activity levels due to their consistent accessibility and operational stability. The analysis of the results reveals that users experience exceptionally high uptime. The computational analysis indicates an average uptime of 99% for large instances (those with over 100 users) and the user survey shows that 97% of respondents perceive their instance as consistently available. Where the descriptive analysis of the variable align in the computational analysis and the user survey, the correlation analysis presents conflicting results. According to the user survey, user activity is negatively correlated with uptime, with a Spearman coefficient of -0.06 ($p < 0.001$). Conversely, the computational analysis reveals a positive correlation, with a Spearman coefficient of 0.25 ($p < 0.001$). This discrepancy might be explained by differences in the perspectives captured by each method. More active users may be more likely to notice instances being not available, leading to a perceived negative correlation in the survey results. In contrast, the computational analysis, which examined thousands of instances, suggests that lower uptime could prevent users from accessing their instances altogether, resulting in a positive correlation with activity rates. Additionally, it is important to note a distinction between the metrics used: the computational analysis considers monthly active and non-active users, while the user survey captures timely patterns of activity. This indicates that while higher uptime generally supports greater activity rates across large instances, more active users may subjectively experience greater disruptions due to downtime.

- **H2b**: Instances with lower latency demonstrate higher user activity levels.

Studies show that younger generations (ages 16–25) prioritize response time when selecting their social networking site applications [51]. This study evaluates how latency affects user activity across all age groups on Mastodon. However, the results indicate a weak or statistically insignificant relationship between latency and user activity. The user survey revealed a Spearman correlation coefficient of -0.07 with a statistically significant p-value (< 0.001), suggesting that higher latency is associated with slightly higher user activity. This counter-intuitive result may reflect that users with higher activity levels are more likely to notice delays compared to less active users. Additionally, only 9% of survey respondents reported experiencing latency on their instance, further suggesting that latency is not a major issue for most users. The computational analysis on the other hand side showed a non-significant relationship between latency on instances and user activity rate. As a result, the correlation between latency and user activity should be interpreted with caution, given the weak association and the limited perception of latency among users.

- **H2c**: Instances with enhanced security measures exhibit higher user activity levels.

The widespread adoption of decentralised social media platforms brings increased concerns about security and privacy risks [33]. The level of security measures implemented on each instance often depends on available financial and technical resources, leading to variations in security across different instances. To assess the impact of security on user behaviour, the user survey examined how important security implementations were to participants, while the computational analysis evaluated actual user activity on instances with stronger security measures. Specifically, the computational analysis considered whether an instance had DNSSEC enabled. The results showed that 99.9% of instances had a valid SSL certificate, while only 21% of large instances had DNSSEC implemented. Given the near-universal adoption of SSL, only DNSSEC was included in the correlation analysis. Surprisingly, the analysis found a negative correlation between DNSSEC implementation and activity rate, with a Spearman coefficient of -0.07 ($p < 0.05$). This suggests that instances with DNSSEC enabled tended to have slightly lower activity levels. In contrast, the user survey revealed that 75% of respondents consider security implementations important. However, the correlation between user-perceived importance of security and actual activity was weakly positive, with a Spearman coefficient of 0.05 ($p < 0.05$).

These findings suggest a discrepancy between users' stated preference for security and their actual behaviour on instances with enhanced security measures like DNSSEC. While security is considered important, other factors—such may play a more significant role in determining user activity. For instance, larger instances generally have more resources than smaller ones, allowing them to implement advanced security measures. However, they also tend to exhibit lower user activity levels overall as will be explained in the instance size chapter below. This raises the question of whether security measures themselves influence user activity or if larger instances with lower activity levels just present better security implementations. The relationship between security and user activity may therefore be intertwined with instance size, making it a crucial factor to consider. These findings highlight the complexity of security's role in decentralised networks and underscore the need for further research into how security perceptions shape user behaviour.

- **H2d**: Instances with greater transparency, including clear privacy policies and terms of service, foster higher user activity levels.

Users often unknowingly disclose private information on social media platforms, which is then stored on centralised servers owned by large commercial companies [62]. As a result, privacy concerns have become one of the most pressing issues associated with centralised social networking sites [45]. In response, decentralised social media platforms have emerged, offering users greater control over their data and technical infrastructure, thus enhancing personal privacy [31]. On Mastodon, instances can choose to disclose their privacy policies and terms of service by linking them on their profile, providing users with a clearer understanding of data handling practices. To examine the role of transparency in user activity, the computational analysis assessed whether instances had privacy policies and terms of service publicly linked, while the user survey explored users' familiarity with these policies and their perceived importance.

Both analyses indicate that transparency is valued by users and that the presence of privacy policies and terms of service is positively correlated with user activity. The user survey revealed a Spearman coefficient of 0.14 ($p < 0.001$), while the computational analysis yielded a slightly stronger correlation with a Spearman coefficient of 0.17 ($p < 0.001$). These findings suggest that transparency regarding privacy policies and Terms of Service plays a meaningful role in user activity on Mastodon. The positive correlation between the presence of these policies and user activity indicates that users may be more likely to engage with instances where they feel informed and assured about how their data is handled. This aligns with prior research on transparency in decentralised online social networks, which highlights that privacy concerns drive users to seek platforms that offer stronger data protection [7]. While the correlation is statistically significant, the relatively modest coefficients suggest that privacy policies are just one of many factors influencing user activity. Further research could explore, whether user actively read and understand the privacy policies or if their mere presence already generates security for users.

- **H2e**: Information overload leads to lower user activity levels.

When users experience information overload, they may feel a sense of regret, which can lead them to discontinue using a social media platform [44]. Applying this to Mastodon, it was demonstrated that it is not a widespread issue. Only 15% of participants reported experiencing information overload, while 20% felt neutral, and the majority (65%) disagreed with the statement. One possible explanation for this is that advertisements often contribute to information overload [90]. Mastodon with its decentralised nature does not produce any advertisements or algorithmic content curation. The correlation analysis found no statistically significant relationship between user activity and information overload, meaning the H2e hypothesis is not supported. This suggests that information overload does not play a meaningful role in influencing user activity on Mastodon. This suggests that most likely due to the low presence of Information Overload, user activity is not influence. Mastodon users can actively choose whom or which topics they want to follow, allowing them to curate their own feed rather than relying on algorithmic recommendations. This might be one factor that eliminates or reduces information overload. Additionally, users who prefer decentralised platforms may already have a higher tolerance for managing their own feeds, reducing the likelihood of feeling overwhelmed. Future studies could investigate the causes of information overload and compare its prevalence on decentralised and centralised platforms. While centralised platforms rely on algorithmic content curation and advertisements, which may contribute to overload, decentralised platforms like Mastodon allow users to control their own feeds, potentially reducing this effect.

- **H2f**: Environmentally sustainable instances have higher user activity levels.

The hypothesis proposed that environmentally sustainable instances would have higher user activity. This was tested through both a user survey and an API analysis, which differentiated between instances hosted on green data centres and those ones that are not. While the majority of survey participants (55%) indicated that sustainability on their instance is important to them, this preference was not reflected in actual user activity. The correlation analysis found no statistically significant relationship in either the computational analysis or the user survey, indicating that the H2f hypothesis is not supported. These findings suggest that while users may express a preference for sustainability, this does not necessarily translate into changes in their platform activity. This could indicate that other factors have a greater influence on user activity than environmental considerations. Moreover, users may choose a sustainably hosted instance based on their values, but this choice does not appear to impact their daily activity levels.

- **H2g**: Smaller instances exhibit higher user activity levels.

Hypothesis H2g was supported by existing research, which suggests that larger instances tend to exhibit lower user engagement than smaller ones [39]. Preliminary analysis also indicated that countries with larger user bases tend to have lower activity rates. The moderator of Mastodon.at even suggests that an optimal community size for engagement falls between 500 and 1,000 users [94]. The correlation analysis from both the user survey and the API data confirmed a negative relationship between instance size and user activity. The user survey yielded a Spearman coefficient of -0.06 (p-value < 0.05), while the API data showed a much stronger negative correlation, with a Spearman coefficient of -0.64 (p-value < 0.001).

These findings suggest that smaller instances may foster a stronger sense of community and engagement among users, while larger instances reduce overall interaction levels. The stark contrast between the relatively weak correlation in the user survey and the strong negative correlation in the API data might be due to differences in how user activity is measured. The survey captures user activity based on self-reported time spent on the platform, whereas the computational analysis calculates an activity rate by dividing the number of monthly active users by the total number of registered users. This discrepancy suggests that while users may not perceive a direct impact of instance size on their daily engagement, the actual participation rates indicate that larger instances tend to have a higher proportion of inactive users. This could be due to smaller communities facilitating more meaningful interactions, while larger instances may suffer from weaker social ties. This interpretation aligns with findings from the open-ended question about other influencing factors, where many users emphasized the importance of community and personal connections as key factors influencing their activity on Mastodon.

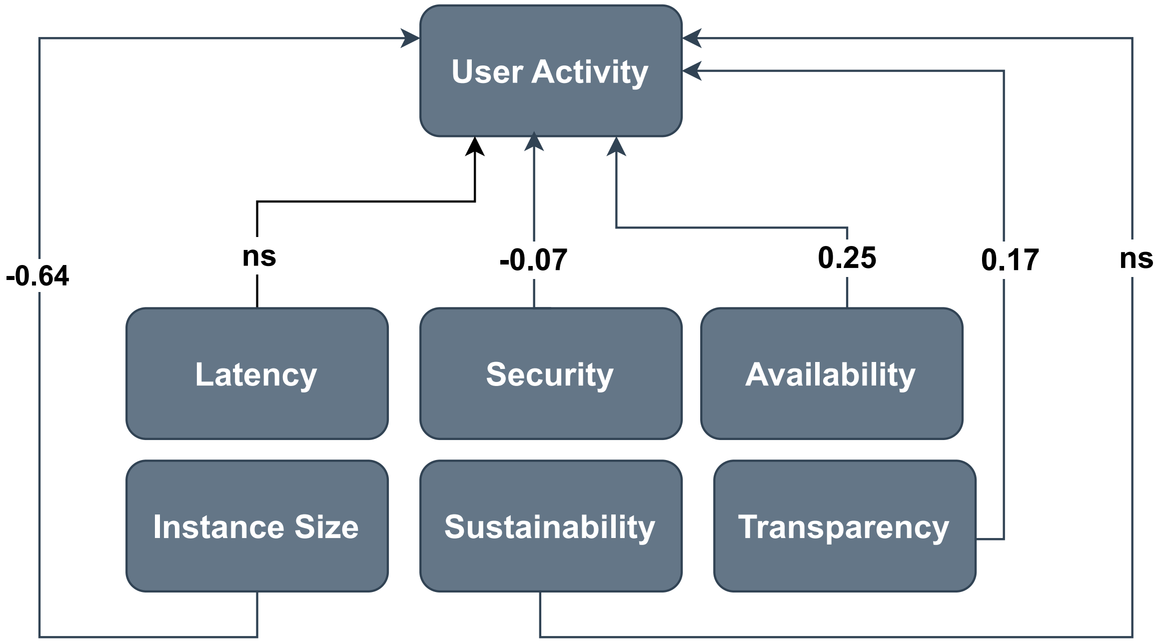

Figure 42: Overview of computational analysis results and the correlation between technical factors and user activity. This figure illustrates the impact of key technical factors on large instances, including instance size, latency, uptime, transparency, sustainability, and security implementation, on user activity levels. The analysis is based on API-derived data and the Spearman correlation analysis.

The third research question was posed as follows:

- **RQ3**: How does active user engagement influence user activity on Mastodon?

This research question focused on whether increased engagement corresponded with higher user activity levels. To investigate this, data was collected through a user survey. Engagement was categorised into three types: contribution, inbound, and outbound engagement, with a separate hypothesis formulated for each. It was hypothesised that all forms of engagement would positively correlate with user activity.

- **H3a**: Users who proactively toot demonstrate higher levels of user activity.

This hypothesis specifically explores the correlation between posting frequency and user activity. The survey results indicate that 38% of participants toot often or very often, 35% toot sometimes, 22% rarely, and 5% never. The correlation analysis revealed a Spearman correlation coefficient of 0.38 ($p<0.001$), which indicates a moderate positive relationship between posting frequency and user activity. These findings suggest that users who actively contribute by posting on Mastodon tend to have higher overall activity levels on the platform. This could be due to the fact that posting often leads users to engage further - replying to comments, reviewing interactions, creating or thinking about their next toot - which naturally increases the amount of time they spend on the platform. Frequent posters are likely to be more engaged in discussions, which keeps them coming back to the platform. However, the correlation also raises questions about causality - whether increased posting drives user activity, or whether already active users are simply more inclined to post. Future research could explore how different engagement patterns, such as passive scrolling versus active posting, influence long-term retention and user activity.

- **H3b**: Users who engage with the content of other users exhibit higher levels of user activity.

Centralised social media platforms rely on algorithmic curation to generate endless feeds, prioritising passive consumption over active engagement [64]. However, the findings of this study suggest that fostering active participation may be a more effective strategy for sustaining user activity. The analysis revealed a moderate positive correlation between user engagement (replies, boosts, and favourites) and user activity. The Spearman rank correlation analysis produced a statistically significant coefficient of 0.40 ($p < 0.001$). These findings suggest that active engagement, such as replying, boosting, and favouring, plays a crucial role in sustaining user activity on Mastodon. Unlike passive scrolling, which is often encouraged by algorithm-driven platforms, interacting with the content of other users appear to reinforce continued platform use. This implies that fostering engagement through social interactions may be a more effective strategy for maintaining user retention in decentralised social media environments. Additionally, for individual users looking to spend more time on Mastodon, this could mean finding content or a community they enjoy engaging with. This underscores the significance of community-driven interactions in shaping user behaviour and fostering sustained engagement.

- **H3c**: Users who receive higher engagement on their toots exhibit higher levels of user activity.

User engagement, including likes, comments, and shares, plays a crucial role in shaping user behaviour on social media. Receiving no feedback may cause disengagement or even platform abandonment [21]. A more recent study found that low interaction on posts is linked to stress, negative emotions, and decreased self-esteem [84]. Additionally, engagement is essential for sustaining professional discussions on social networking sites [85]. In this study, 63% of participants reported receiving some level of engagement when they toot. The findings support the H3c hypothesis, showing a positive Spearman correlation coefficient of 0.19 ($p < 0.001$). This suggests that users who receive engagement on their posts tend to spend more time on the platform. This aligns with prior research suggesting that engagement plays a key role in motivating users to contribute and revisit online communities [67]. However, given the relatively weak correlation, engagement alone may not be the primary factor driving user activity. Furthermore, as this study relies on self-reported perceptions of engagement rather than objective interaction data, future research should investigate the causal relationship between engagement and user activity using quantitative metrics of actual interactions.

12 Limitations

While this study provides valuable insights into the factors correlating with user activity on Mastodon, several limitations must be acknowledged. These limitations arise from methodological constraints, platform-specific considerations and the challenge of interpreting the impact of certain factors on user activity.

12.1 Methodological Limitations

This study employed two primary data collection methods: a user survey and a computational analysis based on API data. While both approaches provided valuable insights, they also come with inherent limitations.

12.1.1 Limitations of the User Survey

Self-reported data bias The user survey was based on the perception of users, which introduces potential biases and inaccuracies. When estimating their own user activity or posting frequency, respondents had to rely on recall rather than actual recorded data. This subjective perception may lead to inconsistencies. For instance, one user may interpret tooting "often" as four to five times a week, while another may consider it to be four to five times a day. Consequently, user behaviour as reported in the survey may not align with actual usage patterns. This presents challenges in standardization and comparability across all the respondents.

Challenges in measuring content moderation Another challenge in this study was measuring content moderation, as it relied on user perception rather than objective metrics. Many users may not directly observe the work of content moderators, making it difficult to assess whether moderation efforts were actively enforcing community guidelines. Additionally, users' exposure to inappropriate content in their feed may not necessarily reflect the efficiency of moderation but rather the timing of their activity—whether they happened to see a post before it was removed. Given these limitations, relying solely on user perception may not provide a comprehensive understanding of moderation effectiveness.

Influence of administrators and moderators The study was initially designed to focus on users of Mastodon instances rather than administrators or moderators, as the goal was to provide insights into how admins and moderators can shape their instances to foster an ideal user environment. However, during the study, a significant number of respondents identified as administrators, moderators, or users of single-user instances. Some of these participants expressed concerns that their responses regarding community guidelines and content moderation could skew the results. In response to this feedback, a filter question was introduced, allowing administrators, moderators, and single-instance users to skip these questions. Despite this adjustment, there remains a possibility of slight bias in the results, as some responses may still reflect perspectives outside the intended user focus. However, the option to select "I don't know" was available, which may have mitigated some of this effect.

Sampling bias Another potential limitation is the representativeness of the sample. The survey was distributed on Mastodon, attracting over 4,000 respondents. However, its reach may have been limited to specific user circles, particularly those willing to participate in research or community discussions. As a result, the sample may not fully reflect the broader Mastodon user base, particularly those who are less engaged with research-related topics.

Focus on active users only The survey focused exclusively on active Mastodon users to identify factors that contribute to increased user activity. However, the study was initially motivated by the significant gap between active and inactive users. Understanding the reasons why users leave the platform would have provided valuable insights, but reaching inactive users presents a considerable challenge. Future studies could develop strategies to reach former users and investigate the reasons behind their disengagement.

12.1.2 Limitations of the Computational Analysis

Potential inaccuracies The computational analysis relied on publicly available API data from Fediverse Observer, which was crawled from Mastodon instances. While this approach allowed for objective measurements, it also had several constraints. For instance, key parameters such as instance size were self-reported by instance owners, which may have led to inaccuracies. This was evident in some outliers, where certain instances recorded implausibly high numbers of registered users—some reporting millions, despite the largest Mastodon instance to date, mastodon.social, having approximately 2.5 million registered users.

Limited range of measurable factors Additionally, the dataset obtained from Fediverse Observer was limited to the factors provided by the API, primarily technical aspects such as instance size, security implementations, and uptime. This meant that certain aspects, such as information overload, could only be measured through user perception, as they do not have direct computational metrics.

12.2 Platform Scope Limitation

Exclusive focus on Mastodon This study focused exclusively on Mastodon, which is just one application within the broader Fediverse. This choice received some criticism from other Fediverse users. However, Mastodon was selected due to its prominence, large user base, and accessibility of public data, making it a suitable case for analysis within the scope of this research.

Diverse communities Unlike centralised platforms, Mastodon's decentralised nature results in diverse instances with distinct policies, moderation approaches, and user cultures. During the course of the survey, a respondent raised the question of whether different communities (e.g., knitting communities or LGBTQ+ spaces) were considered in the study. This highlights a limitation of the study, as it generalises user experiences across instances without accounting for the unique dynamics of specific communities. As users come from various backgrounds and instances with different social norms, this variability may have influenced the findings. Future research could explore how community-specific factors shape user activity and engagement on Mastodon.

Fluctuations in the Fediverse landscape The broader Fediverse ecosystem is constantly evolving, with new instances emerging and others shutting down. These dynamics may have introduced variability in the results that was not fully accounted for in this study.

12.3 Limitations in Identifying Influencing Factors

Other possible factors While this study identified key technical, governance, and engagement factors correlating with user activity, other variables may also play a role but were not included in the analysis. Notably, responses from the exploratory open-text field highlighted additional potential influences, such as community dynamics, content relevance, and current events. These factors could be valuable for future research to gain a more comprehensive understanding of user activity on Mastodon.

Correlation vs Causality Additionally, this study relied on correlation analysis, which does not establish causality. For instance, the computational analysis of security measures showed that larger instances tend to implement higher security standards while exhibiting lower user activity. However, this does not imply that security measures directly reduce user activity. It is equally possible that lower user activity allows administrators more time and resources to implement advanced security measures or that larger instances have the available resources to implement security measures. Since this study did not examine causal relationships, further research is needed to determine the direction and underlying mechanisms of these associations.

13 Future Works

Although this study provides useful insights into user behaviour on Mastodon, several areas remain unexplored. Future studies can extend these findings by addressing the limitations of this thesis and exploring other elements that influence user activity on Mastodon.

Mastodon vs the whole Fediverse This study focused exclusively on Mastodon, but the broader Fediverse consists of multiple decentralised social networking applications, such as Lemmy, PeerTube, Pleroma, and Misskey. Future research could compare user activity across different Fediverse applications to determine whether similar factors influence user behaviour. Examining differences and similarities between platforms could provide deeper insights into engagement dynamics, governance structures, and community interactions.

Community Specific Behaviour This study generalised user experiences across Mastodon instances without differentiating between various types of communities. However, different communities—such as hobby groups, activist networks, or political circles—may have unique governance styles, moderation policies, and user engagement patterns. Future research could dive more into the characteristics of different communities and whether community norms, governance styles, moderation policies or interactions differ on different instances.

Inactive Users A major motivation for this study was the gap between active and inactive users, yet the survey only examined active Mastodon users and the computational analysis only technical factors. Future research should reach out to inactive users through targeted survey or interviews to understand the reason why they chose to leave Mastodon. Understanding these reasons could provide valuable insights into user retention and long-term engagement strategies.

Content Moderation in Decentralised Networks Content moderation is a crucial issue on both centralised and decentralised social media platforms. Users who are not directly affected by moderation may rarely notice its impact, while marginalized groups on centralised platforms often face greater restrictions, bans, or silencing [5, 24]. In this study, moderation was difficult to measure, as some users may not have been aware of moderation actions unless they were directly impacted. Future research could investigate moderation practices using alternative approaches such as analysing moderation logs, studying content removal records, engaging with platform moderators, or interviewing users who have recently encountered content restrictions.

Additional Factors Influencing User Activity This study identified various factors correlating with user activity, but responses in the open-text field suggested additional influences, such as community engagement, personal connections or the type of content. Future studies could explore these aspects to provide a more comprehensive understanding of user behaviour on Mastodon. Furthermore, while this study found correlations, it did not establish causality. Longitudinal studies or experimental designs could help determine the actual cause of user behaviour and not only the correlation. For example, experimental research could manipulate engagement mechanisms on test instances to assess their direct effects on user activity.

Long-term Studies on Mastodon The snapshot nature of this study limits its ability to determine long-term behavioural trends. Future research could benefit from longitudinal studies that track user behaviour over time, assessing how changes in Mastodon communities, the underlying code or the external environment (policy changes in centralised social media platforms, media articles, et cetera) might impact user activity.

14 Conclusion

Following the mass migration of Twitter users to Mastodon in 2022, sustaining long-term user engagement has remained a challenge [85]. Over time, the gap between active and inactive users has continued to grow. This thesis aimed to explore ways to sustain Mastodon's active users by identifying and analysing factors that correlate with user activity.

14.1 Methodology

A computational analysis and a user survey were conducted to examine the correlation between various factors and user activity. The study was based on the following three research questions:

- RQ1: How do different governance structures, such as community guidelines and content moderation, impact user activity on Mastodon instances?
- RQ2: How does the technical infrastructure affect user activity on Mastodon instances?
- RQ3: How does active user engagement influence user activity on Mastodon?

To address these questions, the study investigated twelve key factors hypothesized to impact user activity. These factors were categorized into three main areas: governance structures (community guidelines, moderation practices), technical infrastructure (uptime, latency, security implementations, transparency, information overload, sustainability, instance size), and engagement dynamics (contribution, inbound engagement, outbound engagement). To investigate these factors, the study utilised a quantitative approach, combining a user survey with a computational analysis of Mastodon's publicly available API data. The survey collected self-reported insights from active users regarding their perceptions of governance structures, technical infrastructure, and engagement, while the computational analysis objectively measured the technical infrastructure. By examining employing a dual approach, the study sought to provide a comprehensive understanding of what drives user activity on Mastodon and offer insights into how platform administrators and developers can foster a more engaged and sustainable user base.

14.2 Key Findings

The findings indicate several key insights into the correlations between the hypothesised factors and user activity. One of the most significant findings was the strong negative correlation between instance size and user activity. The computational analysis showed a Spearman correlation coefficient of -0.64, indicating that smaller instances tend to have higher activity rates than larger ones. This suggests that users in smaller communities may feel a stronger connections, while larger instances may dilute social interactions, leading to lower overall user activity.

Another major finding was the role of user engagement in sustaining activity. Both active contribution (posting frequency) and outbound engagement (interacting with content of other users) had some of the highest correlations with user activity. The correlation between outbound engagement and user activity had a Spearman coefficient of 0.40, suggesting that users who actively engage with others through replies, boosts, and favourites tend to spend more time on the platform. Similarly, frequent contributors (users who post more often) showed a positive correlation with user activity (0.38), reinforcing the idea that active participation fosters sustained engagement. Receiving engagement on a post, on the other hand, showed to have a smaller correlation with a coefficient of 0.19.

In contrast, most technical infrastructure factors, including latency, security, sustainability, and information overload, showed little to no significant correlation with user activity. However, uptime in larger instances exhibited a positive correlation (0.25) in the computational analysis, which aligns with expectations, as greater instance availability ensures consistent access, enabling more users to log in and engage with the platform regularly. While technical reliability is crucial for maintaining platform functionality, these findings indicate that infrastructure factors alone do not significantly drive user behaviour. Governance structures, such as community guidelines and moderation practices, showed a weak but positive correlation with user activity (0.14 and 0.13, respectively). This suggests that clear, inclusive and transparent rules and active moderation contribute to a more structured environment, which may help retain users, though they are not primary drivers of user activity.

14.3 Implications

On a theoretical level, these findings deepen the understanding of user activity on decentralised social networking sites, highlighting the key role of social interactions in keeping users on the platform. Unlike centralised platforms that rely on algorithmic content curation and passive scrolling, Mastodon's engagement dynamics indicate that active contribution (tooting) and engaging with other content —such as tooting, replying, and favouring content from other users—is a key driver of user activity.

From a practical perspective, these results suggest that moderators and administrators should focus on fostering engagement and maintaining smaller, more interactive communities to sustain user participation. Managing instance size by limiting registrations could help create a smaller, more engaged user base. While technical infrastructure and governance structures help ensure a stable and functional environment, their direct impact on user activity appears to be limited. Administrators should prioritize essential technical and governance measures but recognize that investing excessive time in these areas is unlikely to significantly boost user activity on their instance. Instead, fostering engagement and community interaction may be more effective in sustaining active users.

14.4 Limitations and Future Work

While this study provides valuable insights into user activity on decentralised social media platforms, several limitations must be acknowledged. The reliance on self-reported survey data introduced potential biases, particularly in estimates of posting frequency and user activity. Another limitation in the user survey was that content moderation was challenging to measure, as it depended on user perception rather than objective data. Additionally, the survey only included active users, limiting insights into why some users disengage from Mastodon. The study's exclusive focus on Mastodon also restricts the generalizability of findings to the broader Fediverse, where engagement dynamics may differ. Lastly, while correlations were identified, causality was not established, leaving open questions about the true drivers of user activity.

Given these limitations, future research should take a broader approach by comparing different Fediverse applications to determine whether similar factors influence user activity across decentralised platforms. Moreover, investigating community-specific behaviours could provide deeper insights into how engagement dynamics vary among different user groups. Future studies should seek to interview inactive users to better understand the reasons behind leaving Mastodon and explore retention strategies. The complexities of content moderation in decentralised networks also warrant further investigation using alternative methodologies, such as moderation logs and interviews with affected users. Finally, studies tracking user behaviour over time could offer valuable insights into how external factors, evolving platform policies, and community shifts impact long-term engagement on Mastodon.

14.5 Concluding Remarks

This study deepens the understanding of user behaviour in decentralised social networks by identifying key factors that correlate with user activity on Mastodon. The findings emphasise the role of social interaction, governance structures, and instance size in shaping user participation, while technical infrastructure serves more as a foundation than a driving force. As decentralised platforms continue to evolve as alternatives to mainstream social media, understanding what keeps users on the platform becomes increasingly important for administrators, developers, and researchers. By highlighting what sustains user activity, this research provides valuable insights for fostering strong and connected online communities. However, with the Fediverse constantly changing and decentralised governance bringing its own complexities, further research is needed to explore long-term user retention, evolving engagement patterns, and the broader impact of decentralised social networking in an ever-shifting digital landscape.

References

[1] Vibhor Agarwal, Aravindh Raman, Nishanth Sastry, Ahmed M. Abdelmoniem, Gareth Tyson, and Ignacio Castro. Decentralised moderation for interoperable social networks: A conversation-based approach for pleroma and the fediverse. *Computing Research Repository in arXiv*, abs/2404.03048, 2024.

[2] Jesús C. Aguerri, Fernando Miró-Llinares, and Ana B. Gómez-Bellvís. Consensus on community guidelines: an experimental study on the legitimacy of content removal in social media. *Humanities & social sciences communications*, 10(1):416–11, 2023.

[3] Thomas Aichner, Matthias Grünfelder, Oswin Maurer, and Deni Jegeni. Twenty-five years of social media: A review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, behavior and social networking*, 24(4):215–222, 2021.

[4] Mohammed Ali Al-garadi, Kasturi Dewi Varathan, and Sri Devi Ravana. Cybercrime detection in online communications: The experimental case of cyberbullying detection in the twitter network. *Computers in Human Behavior*, 63:433–443, 2016.

[5] Kendra Albert and Oliver Haimson. Comments to oversight board. `http://dx.doi.org/10.2139/ssrn.4216597`, August 2022. Trans Content Moderation Scholars.

[6] Ishaku Hassan Anaobi, Aravindh Raman, Ignacio Castro, Haris Bin Zia, Damilola Ibosiola, and Gareth Tyson. Will admins cope? decentralized moderation in the fediverse. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 3109–3120, New York, NY, USA, 2023. Association for Computing Machinery.

[7] Alex Auvolat. Making federated networks more distributed. In *2019 38th Symposium on Reliable Distributed Systems (SRDS)*, pages 383–3831, Lyon, France, 2019. IEEE.

[8] Gökhan Aydin. Social media engagement and organic post effectiveness: A roadmap for increasing the effectiveness of social media use in hospitality industry. *Journal of Hospitality Marketing & Management*, 29(1):1–21, 2020.

[9] Christian V. Baccarella, Timm F. Wagner, Jan H. Kietzmann, and Ian P. McCarthy. Social media? it's serious! understanding the dark side of social media. *European Management Journal*, 36(4):431–438, 2018.

[10] Linnea K. Beckett, Flora Lu, and Sheeva Sabati. Beyond inclusion: Cultivating a critical sense of belonging through community-engaged research. *Social sciences (Basel)*, 11(3):132, 2022.

[11] Mandi Bishop. Healthcare social media for consumer informatics. In Margo Edmunds, Christopher Hass, and Erin Holve, editors, *Consumer Informatics and Digital Health: Solutions for Health and Health Care*, pages 61–86. Springer International Publishing, Cham, 2019.

[12] Douglas G. Bonett and Thomas A. Wright. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of organizational behavior*, 36(1):3–15, 2015.

[13] Carlo Alberto Bono, Lucio La Cava, Luca Luceri, and Francesco Pierri. An exploration of decentralized moderation on mastodon. In *Proceedings of the 16th ACM Web Science Conference*, pages 53–58, New York, NY, USA, 2024. ACM.

[14] Danah m. Boyd and Nicole B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[15] Johannes Breuer, Tarek Al Baghal, Luke Sloan, Libby Bishop, Dimitra Kondyli, and Apostolos Linardis. Informed consent for linking survey and social media data - differences between platforms and data types. *IASSIST quarterly*, 45(1), 2021.

[16] Roderick J. Brodie, Ana Ilic, Biljana Juric, and Linda Hollebeek. Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1):105–114, 2013. (1)Thought leadership in brand management(2)Health Marketing.

[17] Markus Bühner. *Einführung in die Test- und Fragebogenkonstruktion*. ps. Psychologie 4326. Pearson, München, 4., korrigierte und erweiterte auflage edition, 2021.

[18] Umit Can and Bilal Alatas. Big social network data and sustainable economic development. *Sustainability*, 9(11), 2017.

[19] Umit Can and Bilal Alatas. A new direction in social network analysis: Online social network analysis problems and applications. *Physica A: Statistical Mechanics and its Applications*, 535:122372, 2019.

[20] Lucio La Cava, Sergio Greco, and Andrea Tagarelli. Discovering the landscape of decentralized online social networks through mastodon. In *Sistemi Evoluti per Basi di Dati*, 2022.

[21] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. How community feedback shapes user behavior. *CoRR*, abs/1405.1429, 2014.

[22] Carl Colglazier, Nathan TeBlunthuis, and Aaron Shaw. The effects of group sanctions on participation and toxicity: Quasi-experimental evidence from the fediverse. In *18th International AAAI Conference on Web and Social Media*, 2024.

[23] Anwitaman Datta, Sonja Buchegger, Le-Hung Vu, Thorsten Strufe, and Krzysztof Rzadca. Decentralized online social networks. In Borko Furht, editor, *Handbook of Social Network Technologies and Applications*, pages 349–378. Springer US, New York, NY, 2010.

[24] Amber Davisson and Kiernan Alati. "difficult to just exist": Social media platform community guidelines and the free speech rights of sex workers. *Social media + society*, 10(1), 2024.

[25] Statista Research Department. Social media advertising spending worldwide from 2019 to 2029. https://www.statista.com/forecasts/1418549/social-media-ad-spend-worldwide, 2024. Accessed: 31.01.2025.

[26] Stacy Jo Dixon. Number of monthly active facebook users worldwide as of 4th quarter 2023. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/, 2024. Accessed: 02.07.2024.

[27] Stacy Jo Dixon. Social media - statistics & facts. https://www.statista.com/topics/1164/social-networks, 2024. Accessed: 02.07.2024.

[28] Nicola Döring. Publisher erratum zu: Forschungsmethoden und evaluation in den sozial- und humanwissenschaften. In *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, pages E1–E2. Springer Berlin Heidelberg, Berlin, Heidelberg, 2023.

[29] Facebook. Fake accounts, August 2024. Accessed: 2024-10-30.

[30] Fedidb. Fediverse user growth. https://fedidb.org/, 2020. Accessed: 03.07.2024.

[31] Robert W. Gehl. The case for alternative social media. *Social Media + Society*, 1(2), 2015.

[32] Green Web Foundation. About green web foundation, 2024. Accessed: 2024-09-25.

[33] Anatoliy Gruzd, Alyssa Saiphoo, and Philip Mai. The lure of decentralized social media: Extending the utaut model for understanding users' adoption of blockchain-based social media. *PloS one*, 19(8):e0308458, 2024.

[34] Barbara Guidi, Marco Conti, Andrea Passarella, and Laura Ricci. Managing social contents in decentralized online social networks: A survey. *Online Social Networks and Media*, 7:12–29, 2018.

[35] Barbara Guidi, Andrea Michienzi, and Giulio Rossetti. Towards the dynamic community discovery in decentralized online social networks. *Journal of grid computing*, 17(1):23–44, 2019.

[36] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: Marginalization and moderation gray areas. *Proceedings of the ACM on human-computer interaction*, 5(CSCW2):1–35, 2021.

[37] Abdelbaset S Hamza. Optical wireless data center networks, 2018.

[38] Haris Bin Zia, Aravindh Raman, Ignacio Castro, Ishaku Hassan Anaobi, Emiliano De Cristofaro, Nishanth Sastry, and Gareth Tyson. Toxicity in the decentralized web and the potential for model sharing. *Proceedings of the ACM on measurement and analysis of computing systems*, 6(35):1–25, 2022.

[39] Jiahui He, Haris Bin Zia, Ignacio Castro, Aravindh Raman, Nishanth Sastry, and Gareth Tyson. Flocking to mastodon: Tracking the great twitter migration. In *Proceedings of the 2023 ACM on Internet Measurement Conference*, IMC '23, page 111–123, New York, NY, USA, 2023. Association for Computing Machinery.

[40] Amélie Heldt. Reading between the lines and the numbers: an analysis of the first netzdg reports. *Internet Policy Review*, 2(7), 2019.

[41] Daniel Hickey, Matheus Schmitz, Daniel Fessler, Paul Smaldino, Goran Muric, and Keith Burghardt. Auditing elon musk's impact on hate speech and bots. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 1133–1337, 2023.

[42] Manoel Horta Ribeiro, Shagun Jhaver, Savvas Zannettou, Jeremy Blackburn, Gianluca Stringhini, Emiliano De Cristofaro, and Robert West. Do platform migrations compromise content moderation? evidence from r/the_donald and r/incels. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), October 2021.

[43] Juliane House. Authentic vs elicited data and qualitative vs quantitative research methods in pragmatics: Overcoming two non-fruitful dichotomies. *System (Linköping)*, 75:4–12, 2018.

[44] Ziyue Huang, Prashant Palvia, and Nikhil Mehta. The social media discontinuance model: the trio of dark side, regret, and privacy control. *Behaviour & information technology*, pages 1–29, 2024.

[45] Haiyan Jia and Heng Xu. Measuring individuals' concerns over collective privacy on social networking sites. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(1), 2016.

[46] Faith Karimi. The day facebook went dark. *CNN*, October 2021. Accessed: 2024-08-07.

[47] Yonas Mitike Kassa, Rubén Cuevas, and Ángel Cuevas. A large-scale analysis of facebook's user-base and user engagement growth. *IEEE Access*, 6:78881–78891, 2018.

[48] Hamid Khobzi, Ana Isabel Canhoto, and Mohammad Sadegh Ramezani. Content creators at a crossroads with decentralized social media. *Business horizons*, 2024.

[49] Eric Khun. Mastodon analytics. https://mastodon-analytics.com, 2023. Accessed: 2024-07-02.

[50] Kate Klonick. The new governors: The people, rules, and processes governing online speech. *Harvard law review*, 131(6):1599–1670, 2018.

[51] Orhan Korhan and Melike Ersoy. Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. *Quality & Quantity*, 50(4):1799–1816, July 2016. Published online: 14 June 2015.

[52] Lucio La Cava, Domenico Mandaglio, and Andrea Tagarelli. Polarization in decentralized online social networks. In *Proceedings of the 16th ACM Web Science Conference*, WEBSCI '24, page 48–52, New York, NY, USA, 2024. Association for Computing Machinery.

[53] Ganaele Langlois, McKelvey Fenwick, Greg Elmer, and Werbin Kenneth. Mapping commercial web 2.0 worlds: Towards a new critical ontogenesis. *Fibreculture Journal*, 14, 01 2009.

[54] N. Leavitt. Internet security under attack: The undermining of digital certificates. *Computer (Long Beach, Calif.)*, 44(12):17–20, 2011.

[55] Joyce Yi-Hui Lee, Chin-Sheng Yang, Carol Hsu, and Jhong-Heng Wang. A longitudinal study of leader influence in sustaining an online community. *Information and management*, 56(2):306–316, 2019.

[56] Yang Liu, Zhijie Zhao, Jiaying Wang, and Zeguo Qiu. Consumer engagement on social media: an analysis of brand post characteristic combinations. *JOURNAL OF MARKETING MANAGEMENT*, 40(9-10):704–742, JUN 12 2024.

[57] Mastodon. Mastodon documentation. https://docs.joinmastodon.org/, 2024. Accessed: 25.07.2024.

[58] Mastodon. Statistics. https://api.joinmastodon.org/statistics, 2024. Accessed: 2024-07-02.

[59] Inc. Meta Platforms. Meta reports second quarter 2024 results, July 2024. Quarter ended June 30, 2024.

[60] Trisha Meyer and Agnieszka Vetulani-Cegiel. Transparency as an empty signifier? assessing transparency in eu and platform initiatives on online political advertising and actors. *Policy and internet*, 2024.

[61] Matthew N. Nicholson, Brian C Keegan, and Casey Fiesler. Mastodon rules: Characterizing formal rules on popular mastodon instances. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '23 Companion, page 86–90, New York, NY, USA, 2023. Association for Computing Machinery.

[62] Peiyu Pai and David C. Arnott. User adoption of social networking sites: Eliciting uses and gratifications through a means–end approach. *Computers in Human Behavior*, 29(3):1039–1053, 2013.

[63] Chee Wei Phang, Atrevi Kankanhalli, and Rajiv Sabherwal. Usability and sociability in online communities: A comparative study of knowledge seeking and contribution. *Journal of the Association for Information Systems*, 10(10):721–747, 2009.

[64] Maxwell Poole, Ethan Pancer, Matthew Philp, and Theodore J. Noseworthy. Covid-19 and the decline of active social media engagement. *EUROPEAN JOURNAL OF MARKETING*, 58(2, SI):548–571, FEB 8 2024.

[65] Daniel Possler, Sophie Bruns, and Julia Niemann-Lenz. Data is the new oil—but how do we drill it? pathways to access and acquire large data sets in communication science. *International Journal of Communication*, 13:3894–3911, 2019.

[66] Aravindh Raman, Sagar Joglekar, Emiliano De Cristofaro, Nishanth Sastry, and Gareth Tyson. Challenges in the decentralised web: The mastodon case. In *Proceedings of the Internet Measurement Conference*, IMC '19, page 217–229, New York, NY, USA, 2019. Association for Computing Machinery.

[67] Soumya Ray, Sung Kim, and James Morris. The central role of engagement in online communities. *Information Systems Research*, 25:528–546, 09 2014.

[68] Eugen Rochko. Mastodon 4.2: A new search experience and more! https://blog.joinmastodon.org/2023/09/mastodon-4.2/, September 2023. Accessed: 2024-09-28.

[69] Eugen Rochko. Mastodon annual report 2022, 2023. Mastodon gGmbH.

[70] Yoel Roth and Samantha Lai. Securing federated platforms: Collective risks and responses. *Journal of online trust & safety*, 2(2), 2024.

[71] Matthew J Salganik. *Bit by bit : social research in the digital age*. Princeton University Press, Princeton Oxford, 2018.

[72] Andrea De Salve, Barbara Guidi, and Laura Ricci. Evaluation of structural and temporal properties of ego networks for data availability in dosns. *Mobile networks and applications*, 23(1):155–166, 2018.

[73] Zélia Raposo Santos, Christy M K Cheung, Pedro Simões Coelho, and Paulo Rita. Consumer engagement in social media brand communities: A literature review. *International Journal of Information Management*, 63:102457, 2022.

[74] Michael Schaarwächter. Eine alternative ist kein ersatz: Vorgestellt: das soziales netzwerk mastodon / zunehmende popularität im zusammenhang mit skandalen rund um twitter. *Forum für Bibliothek und Information*, 75(4):172, 2023.

[75] Patrick Schober, Christa Boer, and Lothar A. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018.

[76] Hamidreza Shahbaznezhad, Rebecca Dolan, and Mona Rashidirad. The role of social media content format and platform in users' engagement behavior. *JOURNAL OF INTERACTIVE MARKETING*, 53:47–65, FEB 2021.

[77] Nashid Shahriar, Shihabur Rahman Chowdhury, Reaz Ahmed, Mahfuza Sharmin, Raouf Boutaba, and Bertrand Mathieu. Availability in p2p based online social networks. In *2017 4th International Conference on Networking, Systems and Security (NSysS)*, pages 1–9, 2017.

[78] Stacy Jo Dixon. Number of registered mastodon users worldwide as of march 2023. https://www.statista.com/statistics/1376022/global-registered-mastodon-users/, 2023. Accessed: 03.07.2024.

[79] Stacy Jo Dixon. Daily time spent on social networking by internet users worldwide from 2012 to 2024. https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/, 2024. Accessed: 03.07.2024.

[80] Statista Research Department. X (formerly twitter) - statistics & facts. https://www.statista.com/topics/737/twitter, 2024. Accessed: 02.07.2024.

[81] Keith S. Taber. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in science education (Australasian Science Education Research Association)*, 48(6):1273–1296, 2018.

[82] Mohsen Tavakol and Reg Dennick. Making sense of cronbach's alpha. *International journal of medical education*, 2:53–55, 2011.

[83] Mohsin Ur Rahman, Barbara Guidi, and Fabrizio Baiardi. Blockchain-based access control management for decentralized online social networks. *Journal of Parallel and Distributed Computing*, 144:41–54, 2020.

[84] Angelina Voggenreiter, Sophie Brandt, Fabian Putterer, Andreas Frings, Juergen Pfeffer, Luca Maria Aiello, Yelena Mejova, Oshani Seneviratne, Jun Sun, Steffen Staab, and Sierra Kaiser. The role of likes: How online feedback impacts users' mental health. In *Proceedings of the 16th ACM Web Science Conference*, pages 302–310, New York, NY, USA, 2024. ACM.

[85] Xinyu Wang, Sai Koneru, and Sarah Rajtmajer. The failed migration of academic twitter. *ArXiv*, abs/2406.04005, 2024.

[86] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.

[87] World Wide Web Consortium (W3C). Activity vocabulary. https://www.w3.org/TR/activitystreams-vocabulary/, 2017. Accessed: 02.07.2024.

[88] World Wide Web Consortium (W3C). W3c recommendation 23 may 2017. https://www.w3.org/TR/micropub/, 2017. Accessed: 24.07.2024.

[89] World Wide Web Consortium (W3C). Activitypub: W3c recommendation 23 january 2018. https://www.w3.org/TR/activitypub/, 2018. Accessed: 02.07.2024.

[90] Xing-Zheng Xie and Niann-Chung Tsai. The effects of negative information-related incidents on social media discontinuance intention: Evidence from sem and fsqca. *Telematics and informatics*, 56:101503, 2021.

[91] Dawei Xu, Weiqi Wang, Liehuang Zhu, Jian Zhao, Fudong Wu, and Jiaqi Gao. Cl-bc: A secure data storage model for social networks. *Security and communication networks*, 2022:1–13, 2022.

[92] Yunjie (Calvin) Xu, Yinping Yang, Zengyan Cheng, and John Lim. Retaining and attracting users in social networking services: An empirical investigation of cyber migration. *The journal of strategic information systems*, 23(3):239–253, 2014.

[93] Andrew Zolides. Gender moderation and moderating gender: Sexual content policies in twitch's community guidelines. *New media & society*, 23(10):2999–3015, 2021.

[94] Diana Zulli, Miao Liu, and Robert Gehl. Rethinking the "social" in "social media": Insights into topology, abstraction, and scale on the mastodon social network. *New media & society*, 22(7):1188–1205, 2020.

A Appendix

A.1 Questionnaire

Variablen-Übersicht

Fragebogen-Interne Daten

Im Datensatz finden Sie neben Ihren Fragen folgende zusätzliche Variablen, sofern Sie die entsprechende Option beim Herunterladen des Datensatzes nicht deaktivieren.

CASE Fortlaufende Nummer der Versuchsperson
REF Referenz, falls solch eine im Link zum Fragebogen übergeben wurde
LASTPAGE Nummer der Seite im Fragebogen, die zuletzt bearbeitet und abgeschickt wurde
QUESTNNR Kennung des Fragebogens, der bearbeitet wurde
MODE Information, ob der Fragebogen im Pretest oder durch einen Projektmitarbeiter gestartet wurde
STARTED Zeitpunkt, zu dem der Teilnehmer den Fragebogen aufgerufen hat
FINISHED Information, ob der Fragebogen bis zur letzten Seite ausgefüllt wurde
TIME_001... Zeit, die ein Teilnehmer auf einer Fragebogen-Seite verbracht hat

Bitte beachten Sie, dass Sie die Fragebogen-internen Variablen nicht mit der Funktion value() auslesen können. Für Interview-Nummer und Referenz stehen aber die PHP-Funktionen PHP-Funktion caseNumber() und PHP-Funktion reference() zur Verfügung.

Details über die zusätzlichen Variablen stehen in der Anleitung: Zusätzliche Variablen in der Datenausgabe

Section CG: Community Guidelines

[CG01] Scale (fully labeled)
Community Guidelines Transparency
"How do you feel about the community guidelines on your Mastodon instance?"

CG01_01 are clear.
CG01_02 promote inclusivity.
CG01_03 are easy to understand and accessible.
CG01_04 cover issues that matter the most to me.

1 = fully disagree
2 = mainly disagree
3 = neutral
4 = mainly agree
5 = fully agree
-1 = I don't know.
-9 = Not answered

[CG02] Selection
Community Guidelines Familiarity
"Are you familiar with your instance's community guidelines?"

CG02 Community Guidelines Familiarity

1 = Yes
2 = No
-9 = Not answered

Section DA: Datenschutzverordnung

[DA01] Selection
Datenschutz Filter

DA01 Datenschutz Filter

1 = Yes, I agree.
2 = No, I do not agree.
-9 = Not answered

Section EN: Engagement

[EN01] Horizontal Selection
Filter Engagement
"I create and post my own Toots."

EN01 Filter Engagement
1 = Never
2 = Rarely
3 = Sometimes
4 = Often
5 = Very frequently
-9 = Not answered

[EN02] Scale (fully labeled)
Engagement Häufigkeit
"How often do you engage on Mastodon?"

EN02_02 ... favorite other Toots.
EN02_03 ... boost Toots of other users.
EN02_04 ... reply to other Toots.
1 = Never
2 = Rarely
3 = Sometimes
4 = Often
5 = Very Frequently
-1 = I don't know.
-9 = Not answered

[EN03] Scale (fully labeled)
Engagement on my posts
"How do you feel about the engagement on your Toots?"

EN03_01 When I toot something, I usually get some engagement on it.
EN03_02 I am satisfied with the level of engagement I receive on my toots.
1 = fully disagree
2 = mainly disagree
3 = neutral
4 = mainly agree
5 = fully agree
-1 = I don't
know
-9 = Not answered
EN03_03 I wish there would be more engagement on my toots. (reversed)
1 = fully agree
2 = mainly agree
3 = neutral
4 = mainly disagree
5 = fully disagree
-1 = I don't
know
-9 = Not answered

[EN04] Scale (fully labeled)
Engagement allgemein

EN04_01 I think there should generally be more interaction on Mastodon.
EN04_02 I think there should be more interaction on my local instance.
1 = fully disagree
2 = mainly disagree
3 = neutral
4 = mainly agree
5 = fully agree
-1 = I don't know.
-9 = Not answered

Section IN: Introduction

[IN01] Multiple Choice
Social Media Platforms
"Which social media platforms do you use?"

IN01 Social Media Platforms: Residual option (negative) or number of selected options
Integer
IN01_01 Mastodon
IN01_02 Twitter
IN01_03 Reddit
IN01_04 Facebook
IN01_05 Instagram
IN01_06 TikTok
IN01_07 YouTube
IN01_08 LinkedIn
IN01_10 PeerTube
IN01_11 Pleroma
IN01_12 Bluesky
IN01_09 Other
1 = Not checked
2 = Checked
IN01_09a Other (free text)
Free text

[IN02] Horizontal Selection
User Activity
"How often do you use Mastodon?"

IN02 User Activity
1 = Every day
2 = Several times a week
3 = Once a week
5 = Few times a month
6 = Never
-9 = Not answered

[IN03] Selection
User Activity Daily Amount
"On average, how much time do you spend on Mastodon each day?"

IN03 User Activity Daily Amount
1 = 5 minutes
2 = 10 minutes
8 = 20 minutes
3 = 30 minutes
9 = 45 minutes
4 = 1 hour
5 = 2 hours
6 = 3 hours
10 = more than 3 hours
-9 = Not answered

[IN04] Selection
Years on Mastodon
"When did you first register on Mastodon?"

IN04 Years on Mastodon
1 = Less than 1 year ago (2024)
2 = 1 year ago (2023)
3 = 2 years ago (2022)
4 = 3 years ago (2021)
5 = 4 years ago (2020)
6 = 5 years ago (2019)
7 = 6 years ago (2018)
8 = 7 years ago (2017)
9 = 8 years ago (2016)
-1 = I don't know
-9 = Not answered

Section IO: Information Overload

[IO01] Scale (fully labeled)
Information Overload Content
"How do you feel about the content in your feed?"

IO01_01 ... that there is too much content in my feed to keep up with.
IO01_02 ... overwhelmed by the amount of content in my feed.
IO01_03 ... that the amount of content in my feed is manageable.
1 = strongly disagree
2 = disagree
3 = neither agree nor disagree
4 = agree
5 = strongly agree
-9 = Not answered

Section IT: Technical Infrastructure

[IT01] Scale (fully labeled)
Technical Factors 01
"How do you feel towards the infrastructure of your instance?"

IT01_01 My instance is always available.
IT01_03 The security implementations of my instance are important to me (e.g. DNSSEC certificate, SSL implementation)
IT01_04 I am familiar with the privacy policies of my instance.
IT01_06 The privacy policies of my Instance are important to me.
IT01_05 It is important to me that my instance is powered by green, sustainable energy.
1 = fully disagree
2 = mainly disagree
3 = neutral
4 = mainly agree
5 = fully agree
-1 = I don't know
-9 = Not answered
IT01_02 I sometimes experience delays when trying to connect to my Mastodon instance. (latency problems) (reversed)
1 = fully agree
2 = mainly agree
3 = neutral
4 = mainly disagree
5 = fully disagree
-1 = I don't know
-9 = Not answered

Section MO: Admin

[MO01] Multiple Choice
Filter Admin
"What is your role on your Mastodon instance?"

MO01 Filter Admin: Residual option (negative) or number of selected options
Integer
MO01_02 Administrator
MO01_03 Moderator
MO01_04 Owner of Single User Instance
MO01_05 User
1 = Not checked
2 = Checked

Section MP: Moderation Practices

[MP01] Scale (fully labeled)
Moderation practices
"How do you feel about moderation practices on your instance?"

MP01_01 I often come across content in my feed that I think should be removed by moderators. (reversed)

MP01_04 I believe moderators on my instance should take more action. (reversed)

1 = fully agree
2 = mainly agree
3 = neutral
4 = mainly disagree
5 = fully disagree
-1 = I don't know.
-9 = Not answered

MP01_02 I feel like moderators are taking action to enforce the community guidelines on my instance.

MP01_03 I feel like moderators on my instance reply quickly to reports.

1 = fully disagree
2 = mainly disagree
3 = neutral
4 = mainly agree
5 = fully agree
-1 = I don't know.
-9 = Not answered

Section ST: Statistical Information

[ST01] Selection
Gender
"How do you describe yourself?"

ST01 Gender

1 = Female
2 = Male
3 = Non-binary
4 = Let me specify..
5 = Prefer not to say
-9 = Not answered

ST01_04 Let me specify..

Free text

[ST02] Text Input
Age
"What is your age?"

ST02_01 Age

Free text

ST02_01a Age: I prefer not to say

1 = Not checked
2 = Checked

[ST03] Suggesting Text Input
Country
"What country do you live in?"

ST03 Country
1 = Afghanistan
2 = Akrotiri
3 = Albania
4 = Algeria
5 = American Samoa
6 = Andorra
7 = Angola
8 = Anguilla
9 = Antarctica
10 = Antigua and Barbuda
11 = Argentina
12 = Armenia
13 = Aruba
14 = Ashmore and Cartier Islands
15 = Australia
16 = Austria
17 = Azerbaijan
18 = Bahamas, The
19 = Bahrain
20 = Bangladesh
21 = Barbados
22 = Bassas da India
23 = Belarus
24 = Belgium
25 = Belize
26 = Benin
27 = Bermuda
28 = Bhutan
29 = Bolivia
30 = Bosnia and Herzegovina
31 = Botswana
32 = Bouvet Island
33 = Brazil
34 = British Indian Ocean Territory
35 = British Virgin Islands
36 = Brunei
37 = Bulgaria
38 = Burkina Faso
39 = Burma
40 = Burundi
41 = Cambodia
42 = Cameroon
43 = Canada
44 = Cape Verde
45 = Cayman Islands
46 = Central African Republic
47 = Chad
48 = Chile
49 = China
50 = Christmas Island
51 = Clipperton Island
52 = Cocos (Keeling) Islands
53 = Colombia
54 = Comoros
55 = Congo, Democratic Republic of the
56 = Congo, Republic of the
57 = Cook Islands
58 = Coral Sea Islands
59 = Costa Rica
60 = Cote d'Ivoire
61 = Croatia
62 = Cuba
63 = Cyprus
64 = Czech Republic
65 = Denmark
66 = Dhekelia
67 = Djibouti
68 = Dominica
69 = Dominican Republic
70 = Ecuador
71 = Egypt
72 = El Salvador
73 = Equatorial Guinea
74 = Eritrea
75 = Estonia
76 = Ethiopia
77 = Europa Island
78 = Falkland Islands (Islas Malvinas)
79 = Faroe Islands
80 = Fiji
81 = Finland

82 = France
83 = French Guiana
84 = French Polynesia
85 = French Southern and Antarctic Lands
86 = Gabon
87 = Gambia, The
88 = Gaza Strip
89 = Georgia
90 = Germany
91 = Ghana
92 = Gibraltar
93 = Glorioso Islands
94 = Greece
95 = Greenland
96 = Grenada
97 = Guadeloupe
98 = Guam
99 = Guatemala
100 = Guernsey
[...]
-1 = I prefer not to say
-2 = other text response
-9 = Not answered

ST03s Country (free text)

Free text

[ST04] Text Input
Instance
"Which Mastodon instance are you a member of?"

ST04_01 My mastodon instance

Free text

ST04_01a My mastodon instance: I prefer not to say

1 = Not checked
2 = Checked

[ST05] Selection
Instance Size
"Can you estimate the number of registered users on your instance?"

ST05 Instance Size

10 = Fewer than 100
11 = 100 - 1,000
12 = 1,001 - 10,000
13 = 10,001 - 50,000
14 = 50,001 - 500,000
15 = More than 500,000
-1 = I don't know
-9 = Not answered

[ST06] Text Input
Anmerkungen
"Do you have any other comments or thoughts?"

ST06_01 [01]

Free text

[ST07] Text Input
Factors
"What factors do you think influence your activity on Mastodon?"

ST07_01 [01]

Free text